Fostering Deep Understandings of Emergent Science Concepts

by

Man Su

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2023 by the
Graduate Supervisory Committee:

Michelene T. H. Chi, Chair
Brian C. Nelson
Yi Zheng

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

Integrating agent-based models (ABMs) has been a popular approach for teaching emergent science concepts. However, students continue to find it difficult to explain the emergent process of natural selection. This study adopted an ontological framework–the Pattern, Agents, Interactions, Relations, and Causality (PAIR-C)–to guide the design of learning modules. This pre-posttest experimental study examines the effects of the PAIR-C module versus the Regular module on fostering students' deep understanding of natural selection. Results show that students in the PAIR-C intervention group performed better in answering deep questions assessing the understanding of inter-level causal relationships than those in the Regular control group. Although students in both groups did not show significantly improved abilities in explaining the natural selection process for other contexts or significant differences in their abilities to explain other emergent phenomena, students in the intervention group demonstrated system-thinking perspectives and fewer misconceptions in their expressions compared to the control group. A close analysis of student misconceptions consolidates that the intervention group demonstrated drastically fewer categories and numbers of misconceptions while those in the control group did not show such drastic changes before and after the study. To precisely address misconceptions and further improve students' learning outcomes, Epistemic Network Analysis was adopted to capture students' misconception characteristics by examining the co-occurrences of different misconception categories as well as the relationship between misconceptions and PAIR-C features. The results of student learning outcomes and misconception characteristics collectively provide directions for improving the instructional design of the PAIR-C module. Furthermore,

findings on student engagement levels during learning can also inform future design efforts. Overall, this project sheds light on applying an innovative framework to designing effective learning modules to teach emergent science concepts.

DEDICATION

This dissertation is dedicated to the memory of my grandparents who gave me endless love to face challenges with courage and invaluable wisdom to always inquire with curiosity.

I would also like to dedicate this work to my beloved parents who nurtured me in a free, open, and equal environment that allowed me to explore and develop my interests and identity. I appreciate the trust, respect, and understanding they have shown me at every critical point in my life, which has helped me to make important decisions with confidence and clarity. Their unwavering support has been a constant source of inspiration, strength, and motivation, driving me to pursue my goals and stay true to myself. Thank you, Mom and Dad, for being my lifelong friends and supporters.

Finally, I would like to dedicate this work to Ruofei Li, a person whom I hold dear and who holds me in his heart for the past 11 years. I am grateful for every genuine interaction, your patience and kindness, and for being there through all the good times and the bad. You have enlightened my life in countless ways, and I am blessed to have you by my side.

ACKNOWLEDGMENTS

I would not have reached this point in my doctoral journey without the guidance and mentorship of my dissertation committee members. I would like to express my sincere gratitude to Profs. Michelene Chi, Brian Nelson, and Yi Zheng for their support and encouragement throughout my doctoral studies. Prof. Chi, I am grateful for your guidance and mentorship throughout the past few years. Your dedication to research has ignited my passion for conducting educational studies, inspiring me to always delve deeper and ask more profound questions. Your tireless efforts to teach me the importance of being rigorous and focused have shaped my quality as a researcher. Most importantly, thank you for your support and encouragement during this dissertation work. Prof. Nelson, it has been an absolute pleasure collaborating with you, undoubtedly the highlight of my Ph.D. experience. Your expertise in learning sciences and educational technology design and development has been invaluable, and I have learned so much from you. Additionally, thank you for creating an enriching environment to work alongside a talented team of students and scholars. Prof. Zheng, thank you for mentoring me in designing the study and for your detailed feedback and advice on how to create a valid and reliable test measure. Thank you for your suggestions about collecting interview data and creating protocols. Moreover, I am also deeply grateful for your prompt and insightful responses to my emails and questions. I feel fortunate to have had the opportunity to work alongside someone as knowledgeable and considerate as you.

In addition to my committee, I would like to express my appreciation for the encouragement and guidance from Prof. James Paul Gee. Those deep conversations with you have sparked a profound introspection within me, delving into existential questions

such as the essence of humanity and how to live an examined life amidst the complex world. Additionally, you have inspired me to contemplate my research from both sensational and system-thinking perspectives with a lasting goal of fostering collective intelligence.

I would also like to acknowledge Prof. Teresa Foulger, Prof. Mina Johnson-Glenberg, Prof. Ray Buss, Prof. Elizabeth Gee, and Prof. Janel White-Taylor for their mentorship and support in my Ph.D. journey. Through communication and collaboration with you all, I have discovered novel and creative approaches to conducting research and teaching, as well as gain a deeper appreciation for the value and significance of interdisciplinary studies.

I would like to express my gratitude for the wonderful friendships and partnerships that have blossomed throughout this journey. I would like to extend special thanks to Dr. Jesse Ha, Xuesong Cang, and Yue Xin for their support of this dissertation project. I would also like to acknowledge Danielle, Qing, Eva, Jianlin, Blanche, Jeff, Lu, Ye, Yue, Ziyi, Minhao, Zhichao, Weijia, Kailun, Yingxiao, Nicole, Brandon, Cathy, Yohan, Soojeong, Jeremy, Lin, Luis, Taylor, Francheska, Janice, and Kristina for their support and companionship during my Ph.D. program.

Finally, I would like to express my gratitude to the instructional team of the EDT 180 course and all participating students. This dissertation project would have been impossible if you had not been willing to experience the integration of simulation modules into online courses. Your support and dedication have meant more to me than words can express.

TABLE OF CONTENTS

Page

CHAPTER

CHAPTER Page

CHAPTER Page

Page

APPENDIX

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Background and Objectives

For decades, the science education community has called for a deep understanding of science concepts (NRC, 1996; 2012) and has emphasized the ability to explain the causal mechanisms underlying science concepts (National Academics, 2009). However, research shows that students at all levels lack deep understanding and hold robust misconceptions about one particular type of science concept - the emergent process concepts, such as evolution, diffusion, and heat transfer. Although many instructional interventions (e.g., using agent-based models (ABMs) to simulate emergent phenomena) were implemented to foster students learning of emergent process concepts, the degree to which learning was improved remained uncertain (Wilensky & Reisman, 2006; Dickes et al., 2016). Students continue to have misconceptions, especially when explaining emergent causality (Basu et al., 2015; Chi et al., 2012a; Chi, 2013; Peel, 2019; Su et al., 2020; Su et al., 2021; Yoon et al., 2018).

To this end, this study has three primary objectives. The first is to employ an innovative framework–the Pattern, Agents, Interactions, Relations, and Causality (PAIR-C)–to guide the design of the online learning modules. The second is to investigate the effect of the PAIR-C module versus the Regular module on fostering students' deep understanding of emergent process concepts and knowledge transfer. The third is to examine and compare the characteristics of student misconceptions that are expressed through their causal explanations of emergent process concepts.

Furthermore, the study also aims to understand student cognitive engagement with online modules, explore the relationship between cognitive engagement and learning effectiveness, and thereby draw implications for designing interactive online modules for students of diverse backgrounds.

Chapter 1 begins with an overview of the key terms as defined in the literature and used throughout the study. It is followed by (a) the problem statement that points out the research gap this study intends to fill in; (b) an exposition of the research questions and hypotheses; (c) an overview of the rationale and significance of the study, (d) and a roadmap for navigating the remaining chapters.

Key Terms

Emergence

Emergence can be described as a phenomenon or process where numerous individual elements or agents at the micro-level independently interact in some uniform way, which gives rise to complex behavior or non-predictable pattern at the macro-level (Holland, 2000; Wilensky & Resnick, 1999). Many natural and social phenomena are emergent phenomena (Wilensky & Jacobson, 2014).

Chi (2005) refers to emergent phenomena as non-linear complex dynamic processes and uses the term emergent processes to describe them. In *emergent processes*, as numerous individual agents uniformly and independently interact, complex behavior emerges at pattern level[1] that is not exhibited at agent level[2] (Chi et al., 2012b; Holland,

---

[1] In the context of science processes, the pattern level is also referred to as the macro level.

[2] In the context of science processes, the agent level is also referred to as the micro level.

2000). Moreover, this emergent pattern will not arise if the various parts are simply coexisting; there must be interactions of these elements amongst themselves to bring about the higher order pattern.

Agent-based Models

Agent-based models (ABMs) are computer simulations used to study the interactions between individual agents or computational objects and how the interactions give rise to emergent phenomena (Basu et al., 2015; Klopfer et al., 2005; Wilensky & Rand, 2015; Wilensky & Resnick, 1999). ABMs differ from traditional models that often assume linear and additive relationships among components. Within ABMs, agents are programmed to interact with other agents and the environment in specific ways. These interactions often produce unpredictable outcomes.

Researchers posit that when students interact with ABMs, they not only engage in “agent-level thinking” (i.e., thinking about the behaviors of individual agents) but also are given opportunities to explore how pattern-level behaviors emerge from agent-level interactions (Dickes & Sengupta, 2013; Goldstone & Wilensky, 2008; Wilensky & Reisman, 2006). However, researchers also recognize that it is difficult for students to develop a deep understanding of *Emergence* if they only interact with ABMs by adjusting model parameters and explaining the consequential results from manipulating the simulations (Chi et al., 2012a; Dickes et al., 2016; Stroup & Wilensky, 2014).

Inter-level Causal Relationship

Developing a deep understanding of *Emergence* requires being able to make connections between the agent-level behaviors and the pattern-level behaviors, and more importantly, provide correct explanations of *Inter-level Causal Relationships* (ICR).

Within the term “ICR”, “level” refers to the scale of a phenomenon. Researchers identified two description levels and two associated forms of reasoning from students’ explanations of emergent processes (Wilensky & Resnick, 1999). The two levels are the micro and macro levels: the micro level involves the behavior of individuals or agents, which are often smaller in scale or even invisible to our naked eyes, whereas the macro level relates to the group properties, which are at a larger scale with observable patterns. The two forms of reasoning are agent-based and aggregate reasoning which are complementary to each other when describing emergent phenomena. Agent-based reasoning is typically expressed in terms of rules or actions for individual behavior. Aggregate reasoning is expressed in terms of group properties, populations, or rates of change of a pattern (Levy & Wilensky, 2008). Hence, “inter-level” refers to the relationship between micro-level agents and macro-level patterns.

Correct explanations of ICR not only build upon the two complementary forms of reasoning but also involve causal explanations of how the agent-level interactions or micro-level behaviors give rise to the macro-level patterns or outcomes (Chi, 2005; Chi, 2012b; Penner, 2000; Wilensky & Resnick, 1999). Therefore, ICR can also be referred to as agent-pattern causal relationships which are often misconceived by students. Compared to achieving a deep understanding of inter-level or agent-pattern causal relationships, understanding input-consequence causal relationships is less challenging.

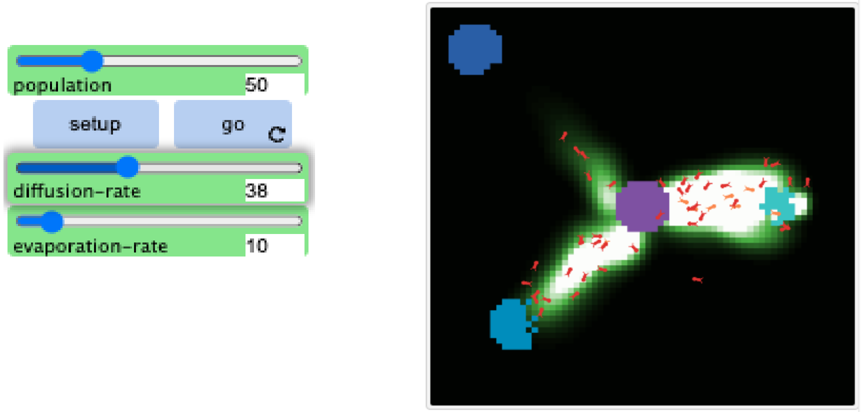
Input-consequence Causal Relationship

Understanding the input-consequence causal relationship is different from understanding the inter-level or agent-pattern causal relationship. Studies have shown that students can explain input-consequence causal relationship by controlling or adjusting the parameters of ABM simulations and interpreting the consequential results (Danish, 2014; Dickes & Sengupta, 2013; Goldstone & Wilensky, 2008; Levy & Wilensky, 2008).

For instance, when reasoning the emergent phenomenon of ants forming lines to search for food, students can explain the input-consequence causal relationship after interacting with the Ants model (Wilensky, 1997b). According to Figure 1, students can adjust or control three parameters (i.e., the number of ants population, the diffusion rate of ants' pheromones, and the evaporation rate of ants' pheromones) in the model, and then observe the consequential patterns of ants' forming queues.

Figure 1.

The Ants Model



population 50

diffusion-rate 38

evaporation-rate 10

As a result, students who adjust the model parameters tend to develop a shallower, input-consequence causal explanation, such as "when the number of ants population increases, the pattern of ants forming lines in search for food appears and disappears at a faster speed". In contrast, a deep inter-level causal understanding will consider the dynamic interactions at the agent level and explain that the line pattern is produced by averaging all the individual ants' distances and directions (towards and away from the line pattern) within each unit of time and then collectively formulate the line across time (Chi et al., 2012b; Chi et al., under review). Therefore, students must understand how to explain inter-level causal relationships as it manifests deep understandings of emergent process concepts.

Research Problem

A substantial number of studies have explored various approaches toward integrating ABMs into science curricula (Basu et al., 2015; Centola et al., 2000; Danish, 2014; Dickes et al., 2016). Nevertheless, it remains unclear whether ABMs are effective in helping students achieve a deep understanding of emergent process concepts given a lack of controlled experiments. Moreover, researchers have not clearly defined inter-level causal relationships nor emphasized the importance of understanding ICRs in learning emergent process concepts. Instead, many efforts focused on developing agent-aggregate complementary reasoning (e.g., describing emergent phenomena from both agent-based and aggregate reasoning perspectives) or input-consequence causal reasoning (e.g., interpreting the consequential results after adjusting model parameters) – and considered such understandings as "deep" understandings of emergent processes (Dickes et al., 2016; Dickes & Sengupta, 2013; Stroup & Wilensky, 2014).

Therefore, this study aims to fill in the research gap by defining and elaborating inter-level causal relationships for emergent process concepts through learning modules that integrate ABMs. Because the PAIR-C framework identifies a set of features that describe and explain ICRs, it is assumed that students can develop deep understandings when using ABMs with multiple representations of PAIR-C features, especially features that explain emergent causal mechanisms.

Research Questions

In this study, I take the position that although ABMs are helpful tools to support students to understand emergent process concepts from both agents- and aggregate-level perspectives, the path leading to deep understanding needs to be carefully designed. Using ABMs without elaborating on the agents-pattern causal relationships is not likely to result in a deep understanding of emergent process concepts.

Specifically, this study adopts the PAIR-C framework to modify existing ABMs, create multiple external representations (MERs), and embed MERs with direct instruction to illustrate PAIR-C features in online learning modules. Using the emergent process concept of natural selection (N.S.) as the focal concept, this study compares the effect of the PAIR-C modules with the Regular modules on fostering students' deep understanding of inter-level causal relationships when explaining the emergent process of natural selection. In response to the study objectives, there are four research questions that this dissertation project aims to address:

RQ1: How to incorporate PAIR-C features, especially features that can explain the inter-level causal relationship, into online learning modules?

RQ2: What are the effects of the PAIR-C module versus the Regular module on fostering students' deep understanding of Natural Selection and knowledge transfer?

RQ3: What are the different misconception characteristics captured from students who engage in the PAIR-C module versus those in the Regular module when explaining Natural Selection?

RQ4: How did the students from both the PAIR-C group and the Regular group engage in the online learning modules?

For RQ1, I describe the design and development of PAIR-C modules in Chapter 5. For RQ2, the hypothesis is that students in the PAIR-C group can show a deeper understanding of N.S. and are more able to transfer their knowledge into explaining emergent causality for different contexts than those in the Regular group. For RQ3, the hypothesis is that students in the PAIR-C group will demonstrate significantly different misconception characteristics (e.g., fewer categories of common misconceptions, less density of misconceptions) from those in the Regular group. For RQ4, the intention is to understand student overall cognitive engagement in self-directed online simulation-based learning environment instead of examining conditional differences.

Significance of Study

This study contributes to research practices on integrating ABMs into emergent complexity science instruction. The PAIR-C framework can impact how instructors design curricula, create assessment rubrics, and use computer technologies to help foster a deep understanding of the causal mechanisms and knowledge transfer of emergent process concepts. The PAIR-C modules designed in this study can be used to provide an alternative instructional approach and help transform the teaching of complex emergent process concepts into flexible and efficient online modules.

Broadly, the study is also conducive to the field of learning sciences, which underscores the importance of understanding emergence as we live in an increasingly complex and interconnected world with ongoing challenges such as global pandemics and environmental degradation (Elsawah et al., 2021; Yoon, 2021; Wilensky & Jacobson, 2014). Since these challenges are essentially emergent phenomena, it is compelling to facilitate a deep understanding of emergent causal mechanisms to probe these challenging issues and prepare the younger generations with a critical, deep understanding for addressing them.

Roadmap to the Dissertation

The remainder of this dissertation consists of eight chapters.

Chapter 2 elaborates on the PAIR-C framework and its features. It also presents relevant literature. First, I conduct a general review of prior research approaches that integrate ABMs into science instruction. I find three main approaches, namely model-building, participatory modeling, and multiple external representations (MERs) supported approaches. Among them, I justify the use of MER-supported ABM as the main approach for this study. Then, I conduct a systematic review that closely examines and evaluates previous ABM integration efforts by adopting the PAIR-C lens. I point out what PAIR-C features were present or absent in previous studies and reflect on how including or not including such features might affect students' learning outcomes.

Chapter 3 focuses on lessons learned from two online pilot studies. The first pilot study suggested that when the PAIR-C features were integrated into a learning module consisting of MERs generated from ABMs, students demonstrated fewer misconceptions and more efficient learning behaviors compared to those who used a learning module embedded with MERs that were non-ABMs. However, students did not show statistically significant differences in the knowledge tests between groups. The second pilot study did not include any simulations but pointed out important information and caveats for implementing an online study. The two pilot studies collectively provide implications in terms of online learning module design, research design, and measurement design, which were fundamental to this dissertation study.

Chapter 4 describes the research methodology employed in this dissertation study, including (a) the study design and setting, (b) methods of data collection, and (c) methods used for conducting data analyses.

Chapter 5 addresses RQ1 by describing and comparing the design and development of the PAIR-C module with the Regular module. It depicted how the PAIR-C module incorporated the seven PAIR-C features into the instruction using the Peppered Moths simulation scenario. It also showed how the Regular module was designed to incorporate common instructional materials.

Chapters 6 to Chapter 8 address the remaining RQs and research objectives respectively. These three Chapters are organized in the same way. Each Chapter contains a brief introduction and an overview. Then, the introduction is followed by findings and a discussion section.

Finally, Chapter 9 presents the general implications of the study with a conclusion.

CHAPTER 2

THEORETICAL FRAMEWORK & LITERATURE REVIEW

The PAIR-C Framework

Chi and colleagues proposed an ontological framework - the Pattern, Agents, Interactions, Relations, and Causality (PAIR-C) framework - to explain the root causes for misconceptions and support instructions on inter-level causal relationships (under review).

Based upon prior works (Chi et al., 2005; Chi et al., 2012a; Chi et al., 2012b; Chi, 2013), the PAIR-C framework identified five dimensions to describe and distinguish two kinds of science processes: emergent processes vs. sequential processes[3]. **P**attern (Dimension I) represents the overall changes by a process that is often visible and meaningful, **A**gents (Dimension II) are elements that participate in the process which produces the pattern, **I**nteractions (Dimension III) refer to how the agents of the process interact, **R**elations (Dimension IV) compare some agents' interactions with other agents' interactions, **C**ausality (Dimension V) refers to the causal relationship between the agents and the pattern. Among the five dimensions, the **P**attern, **A**gents, and **I**nteractions dimensions (Dimension I - III) can be used to describe both emergent and sequential kinds of processes whereas the **R**elations and **C**ausality dimensions (Dimension IV & V) are used to distinguish the two kinds of processes.

The five-dimension PAIR-C framework encompasses and subsumes all nine of the dimensions specified in other theoretical frameworks to describe process concepts,

---

[3] "Sequential processes" are processes that can be decomposed or reduced into a sequence of events or subevents, which are either cyclical or stage-like (Chi et al., 2012a; Chi et al., 2012b).

including actions, sequence, goals, agents, interactions, pattern, structure, behavior, and function (Goel et al., 2009; Hmelo-Silver & Pfeffer, 2004; Wilensky & Jacobson, 2014).

Unlike other frameworks, PAIR-C also provides an ontological perspective to distinguish two kinds of processes by elaborating two kinds of underlying causal structures: Collective causal structure underlies Emergent processes vs. Individualistic causal structure underlies Sequential processes. Although both kinds of processes are studied in the literature (Goel et al., 2009; Hmelo-Silver & Pfeffer, 2004; Mitchell, 2009), there has been no synergistic effort to show how these two kinds of processes can have different underlying causal structures with opposing features on the same dimensions (Chi et al., under review).

To systematically compare and contrast these two kinds of processes, the PAIR-C framework deduces seven features from the Relations and Causality dimensions (first column, Table 1). Among them, there are four *Interaction features* (Feature 1-4, Table 1) identified from the **R**elations dimension. These four *Interaction features* are often perceptible and can provide learners with visual cues to identify different kinds of processes (Features 1-4, Table 1). In addition to the four *Interaction features*, PAIR-C specifies two *Inter-level features* (Features 5 & 6, Table 1) and two inter-level *implication* features (Features 7a & 7b, Table 1) from the **C**ausality dimension. The literature has used an uncategorized list of features and properties, approximately 10-25 features, to explain science processes (Jacobson, 2001; Yoon et al., 2018). In comparison, PAIR-C is more parsimonious by clustering multiple inferences as variations of the same feature or implications features.

Table 1.

The Seven PAIR-C Features for Distinguishing Two Kinds of Processes

Dimension IV: **R**elations among the interactions		
Interaction Features	Sequential Processes (Individualistic Causal Structure)	Emergent Processes (Collective Causal Structure)
Feature 1	Distinct/Different	Uniform/Same
Feature 2	Restricted	Random
Feature 3	Serially	Simultaneously
Feature 4	Dependent	Independent
	Dimension V: **C**ausal relationship between the Agents-pattern	
Inter-level Features	Sequential Processes	Emergent Processes
Feature 5	Incremental change a. The initial pattern is often a part of the final pattern b. New interactions are added to the initial pattern over time	Converging change a. The initial pattern does not resemble the final pattern b. All interactions approximate the final pattern over time
Feature 6	Cumulative summing a. Adding positive & negative numbers between time b. Chain effect: Adding magnitudes & directions between time units.	Collective summing a. Adding positive & negative numbers within time b. Net effect: Adding magnitudes & directions within each time unit.
Feature 7a (Implication)	Single agent is responsible for the pattern i. controlling agent ii. Special status iii. Intentional iv. Teleological v. direct effect	All agents are responsible for the pattern i. No control agent ii. No special status iii. Unintentional iv. Not teleological v. No direct effect
Feature 7b (Implication)	Align or matching between agents and the pattern	Not align or non-matching between agents and the pattern

Table 1 presents the seven PAIR-C features for distinguishing the two kinds of process concepts, which contain four interaction features and two inter-level features with implications. To illustrate the PAIR-C features for the emergent processes, we can match these features to the emergent behavior of ants foraging for food.

First, when searching for food, each ant is acting in a *uniform* way, which is to walk around, emit, and follow pheromones. Yet, the individual behavior of each ant is not the same, even though they are of the same kind. For instance, one ant may walk to the neighbor on the right side, while another ant walks to the neighbor on the left side. The behavior, nevertheless, is of the same kind, that is walking around while emitting and following pheromones of neighboring ants (Feature 1: have the uniform set of actions, see the third column in Table 1).

Second, the interactions of ants are *random*. That is, the ant can adjust its behavior according to any one of its nearest neighbors. Ants can follow or stop following any other ant (Feature 2: random interactions).

Third, ants can follow other ants without any temporal constraint. That is, the ants do not have to take turns, but each can be adjusting its walking directions *at the same time* (Feature 3: interactions occur simultaneously).

Fourth, the interactions in emergent processes are independent of each other. For example, whether an ant follows another is *independent* of whether two other ants follow each other's pheromones (Feature 4: independent interactions).

To understand the causal relationships between the agents and the pattern, it is the prerequisite to first differentiate the kinds of processes (i.e., emergent or sequential) by identifying the above-mentioned four *interaction features* that are often perceivable.

Then, to explain the inter-level causal relationships, it is crucial to know the *inter-level features*. Unlike researchers who used encompassing terminologies (i.e., “small action-big effects, linear versus nonlinear”), the PAIR-C framework proposed that the two inter-level features with the two implication features could be taught to students explicitly which would help students characterize and explain ICR (Chi, 2005; Chi et al., 2012a; Chi et al., 2012b). Again, these features are elaborated using the emergent process of ants foraging for food as an example.

First, the final pattern of an emergent process is formed through *converging change* which all interactions approximate the final pattern over time. Therefore, *the initial pattern does not resemble the final pattern*. For instance, the initial pattern of ants foraging for food is a random distribution of ants. It does not resemble the final pattern of ants forming a single line (Feature 5: converging change).

Second, an emergent pattern is produced by *adding positive and negative numbers* or by *averaging out the magnitudes and directions* of agents *within each unit of time*. For example, the single-line pattern of ants is computed by averaging all the ants’ distances and directions - towards and away from the line. Therefore, the pattern of ants marching in a single line is the collective outcome that sums all the ants interacting within each unit of time. It is *the proportion* of ants staying on a single line increasing over time, not the absolute number of ants on the single line increasing within each unit of time (Feature 6: collective summing).

Third, some of the attributes that describe inter-level causal relationships for emergent processes (i.e., “no control agent”, “no special status”, “unintentional”, “no direct effect”, and “not align”) are classified as implication features. Inferences generated from correct inter-level causal explanations should reflect the two main implication features (Feature 7a: All agents are responsible for the pattern; Feature 7b: Not aligning or non-matching between agents and the pattern). In this case, the interactions of all the ants are responsible for the single-line pattern that emerges. One cannot attribute a single ant (such as the queen ant) as the controlling agent over the other ants to find food in a single line pattern (Feature 7a-i: no control agent). All ants have equal status concerning their contribution to the line formation (Feature 7a-ii: no special status). Each ant merely follows the pheromones without any intention or purpose of producing the single-line pattern (Feature 7a-iii: unintentional; Feature 7a-iv: teleological). Side-stepping of an ant has no direct effect on the overall pattern of marching in a single line (Feature 7a-v: no direct effect). Not all the ants will remain on the single line all the time; some ants can move away from the line if they smell other pheromones (Feature 7b: not align).

Therefore, a correct understanding of ICR for emergent processes requires students to apply the inter-level features and implications corresponding to the emergent processes to describe and explain causal relationships. However, students often conceive of all processes as direct or sequential processes, giving linear, narrative-like, or individualistic causal explanations. Causal ideas such as “A queen ant has a special status and controls other ants to form a single line pattern” and “Ants walk in certain ways with the purpose of forming a single line” are often found in students’ explanations (Chi et al., 2012a; Chi et al., 2012b). Notice that these ideas generated from understandings of sequential processes have antagonistic features against ideas derived from understandings of emergent processes (i.e., “controlling agent vs. no control agent”, “special status vs. no special status”, “intentional vs. unintentional”, etc.). These misconceived ideas are persistent because students are familiar with common sense kinds of causal processes that are already developed from their everyday encounters with stories and sequential processes (Chi et al., 2012a). In contrast, emergent kinds of processes and their underlying collective causal structures are far less familiar to students. Therefore, to foster students' understanding of emergent complex concepts may need to go through a process of radical conceptual change or ontological shift (Chi, 2005; Chi & Roscoe, 2002; Jacobson, 2001).

The PAIR-C ontological framework provides two coherent underlying causal structures with opposing features for discriminating between two kinds of processes. It specifies that the sequential kind of process results from an "incremental" mechanism of "cumulative summing," whereas the emergent kind of process results from a "converging" mechanism of "collective summing" (Chi et al., under review). It is hypothesized that by grasping the PAIR-C features of emergent processes, students can differentiate emergent processes from sequential processes, and correctly describe and explain inter-level causal relationships for emergent processes. Moreover, the accurate identification of the two kinds of processes is fundamental to systematically reducing common misconceptions. To this end, this study will incorporate all seven PAIR-C features into the design of the PAIR-C module.

Relevant Literature

Approaches to Integrate ABMs into Science Instruction

Agent-based models (ABMs) are effective tools in helping students understand complex science concepts (Wilensky & Reisman, 2006). Researchers have developed

ABM environments (e.g., NetLogo) and investigated various approaches for integrating ABMs into science instruction (Wilensky & Resnick, 1999).

Model-building is one common approach that allows students to participate in a trajectory beginning with exploring pre-built models, then examining or modifying the code of existing models, and finally constructing their own ABMs based on their understanding (Centola et al., 2000; Wilensky & Centola, 2007; Wilensky & Reisman, 2006). In the EACH project (Centola et al., 2000), undergraduate students explored and built ABMs to assess the advantages and disadvantages of selfish behavior and how cooperation could evolve. Wilensky and Reisman (2006) later completed an in-depth study to examine the process of how secondary school students developed a predation model using NetLogo. By analyzing students' verbal discussion during the model-building process, researchers found that students began to think of the subject matter as a multilevel phenomenon (Centola et al., 2000). However, a multilevel perspective does not mean a deep understanding of the causal mechanisms that give rise to a multilevel phenomenon. According to Chi et al. (2012b), when exploring or building models of emergent processes, the *collective summing* causal mechanism is opaque to the students since it is often computed by the model itself rather than constructed by students. Researchers need to provide explicit instruction on the collective summing causal mechanism.

Participatory modeling is another common approach that invites students to first "step into" a model by enacting the role of an agent in it, then "step out of" the model to observe the pattern-level outcomes, followed by inquiry activities on how to use agent-level behaviors to explain the collective, pattern-level behaviors (Dickes et al., 2016;

Wagh & Wilensky, 2013; Wilensky & Novak, 2010). In the BEAGLE project (Wilensky & Novak, 2010), students used both ABMs and participatory simulations to explore core mechanisms of evolution. Results showed that students' verbal discussion of the models could manifest increased abilities to reason about emergent phenomena from both agent-level and aggregate-level perspectives and connect agent-level rules and properties with aggregate-level outcomes (Wilensky & Novak, 2010). However, an increased agent-aggregate complementary explanation does not subsume a deep understanding of the causal mechanisms or inter-level causal relationships. Researchers need to clearly distinguish between agent-aggregate complementary explanations with agent-pattern causal explanations when assessing students' understanding.

Researchers have also taken a hybrid approach of using multiple external representations (MERs) to complement ABMs in explaining emergent phenomena (Basu et al., 2015; Chi et al., 2012b; Su et al., 2021; Su et al., in preparation). This approach can be referred to as the MERs-complemented approach. For instance, Chi and colleagues (2012b) have used micro-and macro-level animations and prompts to scaffold students to notice the interaction features and inter-level attributes of the diffusion process, which significantly enhanced students' understanding of diffusion. However, applying a similar approach, students did not show substantial improvement in understanding natural selection (Su et al., 2021). Given that using MERs may support deep understanding by offering multiple perspectives on the same phenomena (Basu et al., 2015), researchers need to probe further into how to use MERs to complement ABMs in illustrating inter-level causal relationships across different process concepts. The MER-supported ABM approach is thereby selected as the main approach for this study.

In sum, the three approaches mentioned above (model-building, participatory modeling, MERs-complemented ABMs) have not reached a consensus on integrating ABMs into science instruction. There was a lack of systematic effort in instructional design that enhanced students' understanding of inter-level causal relationships when using ABMs. Moreover, in the search for more evidence to validate the effectiveness of using ABM to teach emergent process concepts, several issues emerge from the actual implementation of ABMs in science classrooms. In response to these issues, the literature has also suggested possible solutions.

The first issue relates to the visible contents represented in the ABM simulation environment. Many existing ABM simulations do not depict agents' interactions nor the relationships among agents' interactions (Chi et al., under review). This may limit the functionality of ABMs to merely macro-level simulations which show the overall pattern and outcomes after manipulating certain parameters or conditions whereas the micro-level changes of agents' interactions remain obscure to students (Chi et al., 2012b). Based on PAIR-C, depicting the four perceivable interaction features is important because it can help students identify a process as either emergent or sequential, and serve as a prerequisite for understanding the inter-level causal relationships. Therefore, the four interaction features of the emergent process (i.e., uniform; random; simultaneous; independent) need to be included to visualize the relations among agents' interactions.

The second issue also relates to the visual contents presented in the ABM simulation, but from a cognitive load perspective. Numerous studies have reported that students often find it challenging or distracting to accomplish the simulation task because they fail to process the complex information presented in the simulation interface (Basu et al., 2015; Dickes et al., 2016; Dickes & Sengupta, 2013; Xiang & Passmore, 2015). For example, Dickes & Sengupta (2013) revealed that the complexity of information (i.e., many parameters, graphs) represented in the multi-agent-based simulation (MABM) makes it difficult for learners to focus their attention on specific levels of the phenomenon appropriate to the task at hand. Hence, more deliberate efforts need to be made to carefully introduce conditional parameters as well as provide appropriate scaffolds to help students understand the dynamic information presented in ABMs.

The third issue pertains to how the ABM simulation is used. In many interventions that involve either exploring or building ABMs, the inter-level causal relationship (i.e., how the agent interactions give rise to the pattern) is opaque to students since they are often computed by the simulation system itself rather than discovered or computed by students (Chi et al., 2012b). This may result in failures to effectively use ABMs to support students' deep understanding of the underlying mechanisms of their models and the agent-pattern causal relationships of emergent processes (Xiang & Passmore, 2015). Given that understanding the *input-consequence causal relationship* is different from understanding the *agent-pattern causal relationship* (Chi et al., under review), researchers need to focus on integrating ABMs with the specific goal of teaching students about the agent-pattern causal relationships or inter-level features so that their understanding can be possibly improved and become deeper (Dickes et al., 2016; Chi et al., under review).

The fourth issue derives from the third one and considers how the ABM simulation can be enhanced by using multiple external representations (MERs). Previous research has indicated that using MERs may support a deep understanding of emergence by offering multiple perspectives on the same phenomena (Basu et al., 2015). For instance, Chi and colleagues (2012b) have created videos and animations to complement ABMs in illustrating dynamic processes across different timeframes and biological levels, which significantly enhanced learning of the emergent concept of diffusion and deepened the understanding of the agent-pattern causal relationships for the process of diffusion. Hence, to further strengthen the use of ABM simulation, researchers need to generalize this approach by considering how to use MERs to support ABM instruction on other emergent process concepts. This further justifies why the MER-supported ABM approach is selected in this study.

The fifth issue reveals a lack of rigorous empirical efforts to measure the effect of ABM simulations on students' learning outcomes. Most of the ABM projects are pull-out case studies and cannot provide solid evidence of learning gains (Centola et al., 2000; Wilensky & Novak, 2010; Dickes & Sengupta, 2013; Xiang & Passmore, 2015). Only recently, a study involving a larger sample size directly compared the effectiveness of two ABM units on understanding classic evolutionary phenomena and provided empirical evidence on learning gains (Wagh & Wilensky, 2018). The study found that a greater percentage of students in the EvoBuild condition (i.e., model-building ABM unit) provided more evolutionary explanations as compared to those in the EvoExplore condition (i.e., model-exploration ABM unit). However, this empirical study cannot claim that students in the EvoBuild condition manifested a deep understanding of evolutionary mechanisms because of two reasons.

The first reason is that the two ABM units did not give participants equal opportunities to be constructively engaged with the learning materials. Although the two conditions took an equivalent amount of time to cover the same learning contents, students in the EvoBuild condition were engaged in constructive learning activities where they worked in pairs to build models through debugging and testing. For students in the EvoExplore condition, they were mostly engaged in active learning activities where students worked in pairs to manipulate parameters, observe, and explain resulting changes. Based on the ICAP framework of cognitive engagement (Chi et al., 2018), student pairs who are constructively or interactively engaged in learning activities can naturally produce better learning outcomes. Therefore, to make experimental conditions comparable, researchers need to ensure that equal opportunities for cognitive engagement are provided across conditions.

The second reason is that the study did not assess students' deep understanding and did not capture misconceptions expressed in their causal explanations. Since some ideas, such as the input-consequence causal relationships, are more easily learned and understood than other ideas, such as inter-level causal relationships, one cannot claim that an intervention has succeeded if only the easy ideas are learned whereas misconceptions remain in students' causal framing.

Taking account of all these issues, it is critical to reconsider the approaches to integrating ABMs into science instruction, especially to use ABMs to promote a deep understanding of inter-level casual relationships. More efforts should be dedicated to modifying some of the visual contents presented in ABM simulation, designing MERs that support students to elicit a deeper understanding of agent-pattern causal relationships from interacting with ABMs, as well as adopting more rigorous experimental design to examine the effect of ABMs. Moreover, to assess students' deep understanding of ICR for emergent processes, a more definitive, fine-grained, and systematic coding scheme needs to be developed to analyze students' causal understandings and diagnose misconceptions in their causal framing and explanations.

Using PAIR-C to Evaluate ABM Integration Efforts

In this section, we use the PAIR-C framework as an analytical lens for evaluating previous ABM integration efforts. We also analyze what PAIR-C features were present or absent in previous studies and consider how including/not including such features might affect students' learning outcomes. Table 2 is a summary of the analysis.

Table 2.

Analysis of Studies on Integrating ABMs into Science Curriculum

Study	Setting	Content	Approach	PAIR-C Features	Learning Outcomes
Basu et al., 2015	20 8th grade students	Ecosystem	Explore models with MERs	Not explicitly taught, but indicated random and bidirectional interactions	students were able to develop agent-aggregate complementary explanations
Chi et al., 2012a	19 students (11-14 years old)	Natural selection	Observe models	All features except for "collective summing" and "converging change" were taught	students were able to learn agent- and pattern- level ideas; there was little improvement in understanding the causal mechanism
Dickes et al., 2016	15 3rd grade students	Ecology	Participatory modeling & Explore models	Not included	students developed a progressively refined reasoning from both agent and aggregate levels

Dickes & Sengupta, 2013	10 4th grade students	Natural selection	Explore models	Not included	conceptual growth in agent-aggregate complementarity explanations
Wagh & Wilensky, 2018	149 middle school students	Evolution processes	Build models vs. Explore models	Not included	model builders provided more agent-aggregate complementary explanations than model explorers

Not surprisingly, many studies reviewed here that implemented ABMs have shown positive improvements in students' agent-aggregate complementary explanations regardless of different approaches (Basu et al., 2015; Dickes et al., 2016; Dickes & Sengupta, 2013; Wagh & Wilensky, 2018). For example, in Dickes & Sengupta (2013), after interacting with a Birds & Moths ABM simulation, all students could provide agent-aggregate complementary explanations in which they explained aggregate-level outcomes using the agent perspective. Specifically, in reasoning the aggregate-level outcome of darker moths becoming more common, students state "the dark moth population will go up because they will have babies because they're not eaten." (p.932, Dickes & Sengupta, 2013). However, this causal statement was not necessarily correct even though it explains an aggregate-level outcome (i.e., dark moth population goes up) in terms of agent-level behaviors (i.e., dark moths have babies, dark moths are not being eaten).

Similar statements could be found in students' utterances sampled in other studies (Dickes et al., 2016; Wagh & Wilensky, 2018). According to the PAIR-C framework, these agent-aggregate complementary explanations tend to focus on a subgroup of species (i.e., the dark moths) without considering all interactions between all species at the agent level (i.e., the dark moths, the light moths, and the birds who prey on moths). The causal statement seemed to distinguish dark moths' behaviors from other moths as they have special abilities to have babies and avoid being eaten. Moreover, these statements tend to attribute the aggregate-level outcome as static rather than explain how the agent-level behaviors give rise to the aggregate-level outcome by considering all the interactions among the agents. Therefore, agent-aggregate complementary explanations do not necessarily subsume a correct understanding of the inter-level causal relationship.

Compared to these student explanations, a more sophisticated causal explanation should state "the dark moth population will go up because in most generations dark-colored moths survive from being spotted by birds and those survived can reproduce, compared to light-colored moths who have lower survival and reproduction rates (due to industrial pollution, trees became darker with soot and thereby dark-colored moths could blend in the environment more easily than the light-colored moths). Over many generations, dark-colored moths will become more common." Since previous studies already used models with multiple breeds of agents governed by the predator-prey relationship (e.g., both birds and moths are shown in the ABM simulations), explicit instruction on the *Relations among the agents' interactions* (i.e., having the *same/uniform set* of interactions: all moths can reproduce with each other and can be eaten by birds, see Feature 1, Table 1) should be available to students.

Existing studies also reveal that researchers often used ABMs to teach the simple idea of "interactions between agents" rather than using ABMs to teach the PAIR-C interaction features. There seems to be confusion about understanding the "interactions" with understanding the "interaction features". According to PAIR-C, "interaction features" of the emergent process refers to the four perceivable features (i.e., uniform, random, simultaneous, independent) which are generated from comparing one pair of interactions with another pair of interactions. In contrast, "interactions" (i.e., mating interaction, predator and prey interaction) refers to the relationships between two agents.

Basu et al. (2015) used a Saguaran ecosystem model to teach students about the interactions between agents (e.g., "Doves eat seeds of the cacti", "Rats eat pods of the ironwood trees", "Hawks prey on rats", "Hawks prey on doves", etc.) without mentioning the interaction features, such as whether one interaction can occur at the same time as another interaction is occurring (i.e., Feature 3: simultaneous from Table 1). Their study showed that understanding the interactions between agents did not help students reason in causal chains unless they were scaffolded to notice the *simultaneous and bidirectional* nature of interactions. When asked, "Knowing that hawks eat rats and rats eat pods, what would happen if hawks were removed from the ecosystem?", all students in their study initially stated that "If there were no hawks to eat the rats, rats would increase. So, pods would decrease and disappear soon." What was missing in students' responses was the ability to reason further about the consequences of the lack of pods on the population level of rats (missing Feature 2 from Table 1). To remedy this issue, Basu et al. (2015) later introduced an external representation tool (i.e., the causal map) as a complementary approach to scaffold students' understanding of the bidirectional nature of the food chain causal relationships. By visualizing the bidirectional interactions between pods and rats, hawks and rats using the causal mapping tool, students showed significant improvement in their causal understanding of the Saguaran ecosystem.

In the literature analysis, only one study attempted to teach the PAIR-C features to students through ABM (Chi et al., 2012a). However, Chi acknowledged that although most *Interaction features* and some *inter-level implication features* (i.e., no causal agent, no alignment between agents and the pattern) were taught, the instruction did not elaborate on how the interactions at the agent level produced the changing pattern. Their instructions did not specifically teach ideas about collective summing or proportion change. Due to the lack of instruction on explaining causal mechanisms, results showed that students' learning was restricted to the basic understanding of the definitions, the nature of agent interactions, and pattern-level behaviors when learning a more difficult concept - natural selection. There was little improvement in understanding the mechanism of how agents' interactions cause the observed pattern. Therefore, to facilitate students' deep understanding of emergent process concepts, future works can build on this line of research by explicitly teaching the inter-level features of "converging change" and "collective summing", explaining how the pattern arises from agent interactions.

In addition, one caveat that needs to be noticed is that prompt questions used in many studies tend to ask students to reason about the input-consequence causal relationship between the model parameters and the aggregate graphs rather than asking students to reason about ICR and the causal mechanisms that produce the pattern (Dickes et al., 2016; Dickes & Sengupta, 2013; Wagh & Wilensky, 2018). For example, in Dickes et al. (2016), the instructor scaffolded students in explaining the pattern-level outcomes of the ABM in terms of what agent-level parameters they selected and reasoning what effect each of the variables had in terms of outputs. In another example (Wagh & Wilensky, 2018), students followed a Predict-Run-Manipulate-Run-Explain cycle to understand the aggregate pattern by observing a histogram that showed the frequency of variations while the model run under different conditions. During this model exploration activity, no prompts or instructions were given to direct students' attention to describing or reasoning about the causal mechanisms of the aggregate pattern. Since understanding agents-pattern, inter-level causal relationships are more challenging than understanding linear input-consequential causal relationships (Chi, under review), more deliberate efforts are needed in designing prompt questions to elicit students' reflections on ICR throughout the learning activities.

Overall, Chapter 2 suggests that future works on ABM integration should consider: 1) using the MERs-supported ABM approach to illustrate the PAIR-C features, especially the feature of "converging change" and "collective summing"; 2) designing prompt questions to probe student to reason about ICR and emergent causal mechanisms while exploring the models.

CHAPTER 3

PILOT STUDIES

Before implementing the pilot studies, the PAIR-C project developed and tested a series of interventions to address students’ persistent misconceptions about the cause-effect relations of scientific phenomena. The PAIR-C interventions designed by the team at the Learning and Cognition Lab were notably different from current instructional approaches in that they would address multiple process concepts and phenomena across a variety of disciplines by teaching the causal structure underlying those concepts and phenomena rather than focusing on remediating one misconception at a time.

To achieve the research goal, the team ran a pre-pilot study to determine the most effective module or combination of modules. The team then revised the modules based on the findings and later ran a classroom study to compare the most effective module(s) identified in the pre-pilot study against a business-as-usual control condition. The classroom study investigated whether the PAIR-C framework if taught to students, would help them understand two emergent processes (i.e., diffusion and natural selection) which were part of their school curricula.

Since students’ understanding of emergent processes can be enhanced by viewing and exploring dynamic simulations (Chi et al., 2012b), the team included simulations as part of the intervention. In the classroom study, students from both conditions were provided with an identical Pearson simulation to learn about *Diffusion* and an identical PhET simulation to learn about *Natural Selection*. These two simulations were non-ABMs and were chosen because they were the ones the participating teachers normally used in their classrooms.

The PAIR-C project hypothesized that providing an overarching framework—in contrast to other frameworks often presented to students such as The Nature of Science—might help students understand concepts of diffusion and natural selection better than if they were only given The Nature of Science before the content instruction. The findings showed that the PAIR-C intervention approach continued to be successful for the simple but robustly misconceived science process, *diffusion*. However, the project did not achieve significantly better learning results with the PAIR-C instructional intervention for the second far more difficult concept, *natural selection*. To generalize the PAIR-C intervention as an alternative instructional approach in facilitating the understanding of more complicated concepts, I joined the project team and focused on improving the instruction by designing and developing a separate learning module illustrating the PAIR-C features using MERs. Specifically, I examined the effectiveness of MERs generated from different types of simulations (ABMs vs. non-ABMs) in facilitating students' deep understanding of Natural Selection and implemented two other pilot studies as described below.

Pilot Study 1

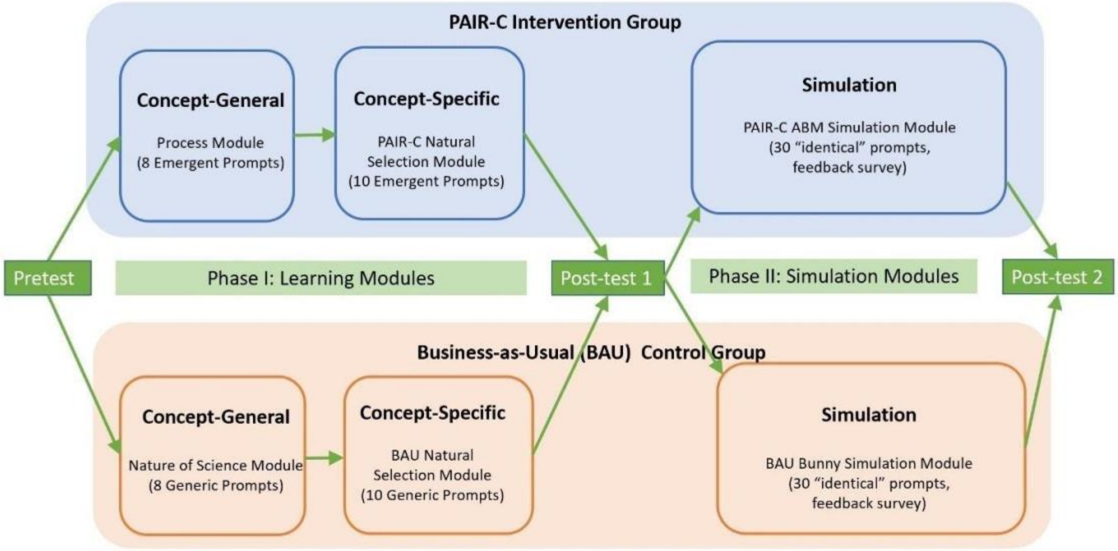
Introduction

Pilot study 1 was conducted to investigate the effect of two kinds of simulation modules on improving students' understanding of the emergent process of natural selection. Students in the Intervention group used the PAIR-C module which was designed to embed MERs generated from ABM simulations whereas students in the Control group used the Business-As-Usual (BAU) module which was designed to embed

MERs based on the commonly used PhET simulation. The primary research question asks: what are the effects of the PAIR-C module versus the BAU module on fostering students' understanding of natural selection? Note that Pilot Study 1 was originally part of the large PAIR-C study which included two other concept learning modules (See Phase I in Figure 2) before the learning module that integrated MERs (See Phase II in Figure 2). Given that the results from the second phase of the study informed my research in designing and investigating the effect of online learning modules consisting of MERs generated from simulations, I treated it as a pilot study for my dissertation.

Figure 2.

A Schematic of the Sequencing of the Instructional Materials Used by the PAIR-C Intervention Group (top) and the BAU Control Group (bottom).



Participants and Settings.

Participants were recruited from two summer non-residential programs. A total of 28 students participated in this online study, including 7 boys and 21 girls, aged 14-18 years old. Among the participants, 79% identified themselves as Asian, with the rest of the group identified as White, Hispanic, or mixed racial or ethnic background. The majority of the participating students were attending high school while 3 participants were college freshmen. Only one student expressed not taking any biology class before the study.

Procedure.

The entire pilot study provided students with 3-4 hours of instruction. Participants were first assigned into two groups using a stratified random assignment approach based on their pretest scores. This resulted in 14 students in the Intervention group and 14 in the Control group whose average pretest scores had no significant difference across groups. As shown in Figure 2, post-test 1 was given at the end of phase one. The second phase focused on the simulation-based module, followed by a feedback survey and post-test 2. Both the PAIR-C and the BAU modules were delivered to students using *Google Forms*. Throughout the study, students learned the instructional materials at their own pace. The pre-posttests were entered into Qualtrics, and students were given links to access the test upon finishing the corresponding module. It was expected for students to complete the second part of the study within an hour.

Design.

In the second phase of the study, both modules were well-structured containing multiple external representations (MERs) in the form of videos, animations, screenshots, tables, and texts to illustrate the emergent process of natural selection. To be specific, the PAIR-C module consisted of MERs modified from the NetLogo wolf-sheep predation ABM (Wilensky, 1997c) to illustrate the four *interaction features* and the three *inter-level features* derived from the PAIR-C framework. In contrast, the BAU module embedded MERs adopted from the PhET Natural Selection simulation (n.d.) to illustrate the parameters, graphs, and charts shown in the simulation interface. Students in both groups interacted with the module mainly by reading instructional texts, playing simulation videos, and typing down responses to the prompt questions. Both modules consisted of seven parts with a total of 30 prompt questions. The only difference between the prompt questions was that the PAIR-C group centered the prompt around sheep whereas the BAU group used bunnies.

Data Sources.

To address the research question in terms of examining the effect of two different simulation modules, the study analyzed the following data: scores from three knowledge assessments (i.e., pretest, post-test1, and post-test2), simulation prompt question responses, feedback survey responses, and time-stamped information from YouTube Analytics to understand student interactions with simulation videos.

Measures.

The knowledge assessment each contained 20 multiple-choice questions which were mostly adapted from conceptual inventories and past AP Biology Exams. Among these items, 12 of them were labeled as hard questions, which asked students to compare agents' interactions and consider the inter-level causal explanations for natural selection. There were also 5 medium and 3 easy questions aimed to assess students' understanding of basic concepts of natural selection. 18 out of 20 questions remained identical across the pre-posttests.

The simulation prompt questions contained 21 binary prompts that asked students "Yes/No" or "True/False" questions; and 9 open-ended prompts that asked students to explain their answers. Some of these prompt questions were two-tiered which the first tier activated students' prior knowledge of natural selection by asking a Yes/No question whereas the second tier prompted students to constructively make predictions or give explanations based on their observations from the simulation module. The prompt questions were designed to not only scaffold students' learning but also elicit their understanding of the PAIR-C features. For example, to elicit students' understanding of the collective summing causal mechanism (Feature 6, Table 1), a first-tier binary prompt asked, "Are white sheep always added to the population between generations to create a pattern of white sheep becoming more common?". Then, it was followed by a second-tier open-ended prompt "Please provide explanations to your above answer in at least 2-3 sentences." Among the 30 prompt questions, six of them were used to elicit an understanding of the inter-level PAIR-C features which focused on testing students' explanations of the causal relationships between agents and the pattern.

The feedback survey contained three five-point Likert scale items that asked students to rate the degree of whether they thought the simulation module was engaging, helpful, or clear. Items ranged from (1) Strongly disagree to (5) Strongly agree. It also included four open-ended questions that asked students to describe the confusing part of the simulation module, give suggestions for visualizing the natural selection process, reflect on their experience when interacting with the simulation videos (i.e., times of replay, speed adjustment, view in full screen or not), and give general comments of the simulation module.

Coding and Scoring.

Because misconceptions sourced from ontological mis-categorization of the emergent process into the sequential process, students' prompt responses were coded based on the seven features listed under the sequential processes (See Table 1). These are opposing features against those listed under the emergent processes.

A two-step process was used to score students' open-ended responses on a 0-2 scale. First, each open-ended response was separated by idea unit and scored for correctness. Second, each idea unit was coded for misconceptions based on the seven opposing features of emergent processes. Two experienced researchers checked the scoring rubric for each open-ended prompt question. Initial agreement among the coders was approximately 90%. Disagreements about misconceptions in the responses were discussed until an agreement was reached.

Findings

Results on Pre-Posttests.

For both groups, participants showed high mean percentage correctness on the pretest (69% correctness for the PAIR-C intervention group; 68% correctness for the BAU control group), indicating that participants were readily able to answer the test based on prior knowledge before the experiment.

Results showed that for students in the PAIR-C group, their performance on the 20 multiple-choice questions in post-test 1 significantly improved $t\ (13) = 3.23$, $p < 0.05$, with a percentage of correctness increased from 69% to 79% and large effect size (Cohen's d = 0.86). For students in the BAU group, their performance on post-test 1 also significantly improved $t\ (13) = 2.85$, $p < 0.05$. The percentage of correctness increased from 68% to 80% correctness with a large effect size (Cohen's d = 0.76).

After completing the simulation module, they took post-test 2 assessing their knowledge of natural selection again. For both groups, no statistically significant improvement was found from post-test 1 to post-test 2. To determine whether there was a difference between the PAIR-C group and the BAU group, ANCOVA tests were conducted and there was no significant difference between the two groups in either post-test 1 or post-test 2 using pretest total score as covariate F (1, 25) = .151, p = .701, partial η^2 = .006; F (1, 25) = .010, p = .923, partial η^2 = .000. Using post-test 1 total score as the covariate, again there was no significant difference found between the two groups F (1, 25) = .532, p = .473, partial η^2 = .021.

Given the interest in examining students' deep understanding of inter-level causal relationships, paired samples t-tests were conducted for 12 hard questions on ICR. For the PAIR-C group, results showed that there was a statistically significant improvement in scores of hard questions from the pretest (M = 8.71, SD = 2.05) to post-test 2 (M = 10.00, SD = 1.41), t (13) = 2.86, $p < 0.05$. However, for the BAU group, there was no statistically significant improvement from the pretest to post-test 2 on the hard questions (M = 9.57 SD = 2.31), t (13) = 1.58, $p > 0.05$. Nevertheless, both Groups did not demonstrate significant improvement from post-test 1 to post-test 2 (p > 0.05) in terms of hard questions. ANCOVA results showed that there were no statistically significant differences found between the two groups, F(1, 25) = .388, p = .539, partial η^2 = .015 when using scores of pretest hard questions as covariates. In sum, although there was a significant increase in students' pre-posttest performance for both groups, no statistically significant differences were found between the two groups.

Results on Prompt Responses.

To further examine the differences between the two groups, we then analyzed the prompt responses within the simulation modules. We conducted ANOVA on the scores of the six prompts assessing students' understanding of causal relationships. The scores for the overall six prompts did not show a significant difference between the two groups F (1,166) = 1.64, $p > 0.05$, $\eta^2 = .01$. This result echoes our findings from the pre-posttests which indicates a lack of differences in students' deeper understanding of inter-level causal relationships for natural selection.

We further examined all the open-ended responses with a score of "1" or "0" in the simulation module and counted the frequency of different types of misconceptions for both groups. Noticeably, there were 40 misconceptions on interpreting Relations among interactions found in the BAU group while there were only seven of them found in the PAIR-C group. Among the 40 misconceptions, more than half of them (N=24) were coded as "restricted". There were also 29 instances of misconceptions for the BAU group when prompted to explain causal relationships. In contrast, 18 misconceptions were found in the PAIR-C group.

In brief, the results showed that students in the PAIR-C group outperformed those in the BAU group in terms of showing fewer misconceptions. The results also indicated that using prompt questions in the simulation module allowed us to detect misconceptions about students' understanding of natural selection between the two groups.

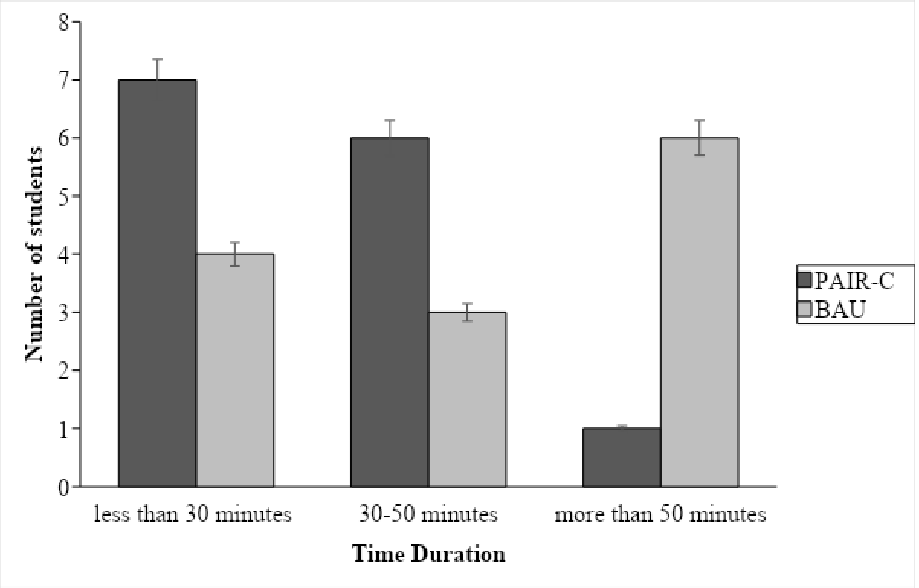
Results on the Feedback Survey & YouTube Analytics.

Students in the PAIR-C group reported an average of 4.14 for engagement and 4.21 for clarity out of 5 in the feedback survey. For the BAU group, the average scores for engagement and clarity were 3.86 and 3.36 respectively. In terms of their learning experience using the simulation videos, eight students from the PAIR-C group reported that they did not re-play any simulation video. However, only two students from the BAU group expressed that they played the simulation videos once. Other students from the BAU group stated that they replayed most videos multiple times.

At the beginning and the end of each simulation module, we also asked students to report their local time. After comparing the time duration (minutes spent in the module) for students to complete the simulation module, we noticed that students in the PAIR-C group spent much less time than those in the BAU group (See Figure 3). Half of the students from the PAIR-C group finished the simulation module in less than 30 minutes while most of the students (9 out of 13) from the BAU group spent more than 30 minutes completing the simulation module. Note that one student from the BAU group did not report the time spent completing the simulation module, therefore the results include 13 responses from the BAU group and 14 responses from the PAIR-C group.

Figure 3.

Time Spent on Two Modules



Results presented from *YouTube Analytics* further corroborated the above findings. Table 3 shows us three instances that compared students' performance in terms of their average view duration and average percentage viewed for a video. These two metrics gave information on students' retention of the simulation videos and how they interacted with certain videos. "Average view duration" indicates the time spent on watching each video, and "Average percentage viewed" provides information on how often each moment of the video is being watched as a percentage of total views for this video. Hence, re-watching certain parts of the video can result in values higher than 100% for the "Average percentage viewed". Take Location C as an example, students in both groups watched a recorded simulation video showing agent interactions for roughly 30 seconds. The simulation video in the BAU group used a pedigree chart to indicate agent interactions and relations whereas the video in the PAIR-C group used "links" to represent random interactions between agents. As shown in Table 3, students in the BAU group watched the simulation video at Location C for an average of one minute and seven seconds, and the average percentage viewed was 249.6%. In contrast, for students in the PAIR-C group, the average view duration of the simulation video at Location C was only 40 seconds with an average percentage viewed of 135.1%. This suggests that students in the BAU group spent more time watching and repeating this simulation video to understand agent interactions than those in the PAIR-C group.

In summary, the findings from Pilot Study 1 suggest that when the PAIR-C features were integrated into a learning module embedded with MERs generated from ABMs, students seemed to demonstrate fewer frequencies of misconceptions about natural selection. However, students did not demonstrate statistically significant differences in the test assessments between groups, their prompt responses during the simulation module provided evidence that also suggests a lack of differences in their understanding of inter-level causal relationships between groups. Other evidence from the study suggests that the PAIR-C module is more time-efficient compared to the BAU module in facilitating self-paced learning in an online environment.

Table 3.

YouTube Analytics

Simulation Videos	Group	Simulation Video Content (length)	Average View Duration	Average Percentage Viewed
Location A	BAU	Wolves-Bunnies (1:49)	1:28	82.4%
	PAIR-C	Wolves-Sheep (1:51)	0:56	51.0%
Location B	BAU	Changing Pattern (1:19)	0:56	72.1%
	PAIR-C	Changing Pattern (1:34)	0:57	61.8%
Location C	BAU	Bunny Pedigree Interactions (0:28)	1:07	249.6%
	PAIR-C	White Sheep Interactions (0:31)	0:40	135.1%

Discussion

Pilot Study 1 was the first attempt to design and implement online PAIR-C modules with MERs generated from ABMs to foster a deep understanding of emergent processes. Although the overall results did not favor the PAIR-C Intervention group over the BAU Control group, the research informed and inspired the dissertation study in multiple aspects.

First is the design and development of the PAIR-C module. The pilot study showed how a regular agent-based model (ABM) could be modified and how multiple external representations (MERs) could be used to instantiate the seven PAIR-C features in the context of natural selection. The pilot study also found that students in the PAIR-C intervention group spent much less time going through the simulation module compared to the BAU control group which indicated that the PAIR-C module might be more effective and efficient in illustrating natural selection as an emergent process compared to the BAU module. However, there were limitations in the design and development process:

1) Only one regular ABM was revised and used as an example to illustrate natural selection throughout the entire module. Robust findings in the cognitive science literature have suggested using two examples is significantly better for understanding than one example (Gick & Holyoak, 1983; Loewenstein et al., 1999). Therefore, the effect of the PAIR-C module might be strengthened by including two modified ABM simulation scenarios to illustrate natural selection.

2) Because the pilot study is part of the large PAIR-C study that introduced the PAIR-C features in two other concept learning modules, the third module was mainly developed for assessing students' understanding of the seven PAIR-C features from their interactions and responses. No explicit instruction or feedback was provided to students that explain how the PAIR-C features could be manifested in the MERs with simulation videos. Given the PAIR-C framework hypothesized that understanding the four perceivable interaction features is the prerequisite for explaining the inter-level causal relationships, future efforts could concentrate on providing explicit instruction and timely feedback to ensure students' understanding of the four interaction features before introducing the more complicated inter-level features. Moreover, when it comes to

teaching the inter-level features, more instruction is needed in explaining how the pattern transforms and how to quantitatively compute the pattern in the context of the simulation scenarios.

3) In the pilot study, we restricted students from directly interacting with the simulation because we wanted to ensure that each participant observed the same version of a random model in a self-paced online learning environment. By only using MERs in the form of recorded videos, animations, and screenshots generated from the modified ABM, we did not provide students with full control of the simulation which might hinder them from probing scientific phenomena and testing their hypotheses. Future works might consider the combined use of an interactive model with MERs. In other words, using MERs as complementary instructional materials to an interactive ABM simulation might further improve students' engagement and the effectiveness of the PAIR-C module.

Second is the pre-posttest study design. Because the pilot study was part of the large PAIR-C project which implemented two learning modules (e.g., the concept-general Process Module, the concept-specific PAIR-C Natural Selection Module) and two tests (e.g., pretest, post-test 1) prior to the simulation module in a short time frame. We could not rule out the carry-over and the retention effect from completing the other learning modules and the repeated knowledge tests. This issue could be resolved in two ways:

1) To further validate the effect and efficiency of using the third PAIR-C module, future studies should conduct a stand-alone study. Instead of comparing the PAIR-C module containing MERs generated from ABMs with the BAU module containing MERs

generated from non-ABMs, future studies should compare the PAIR-C module with the Regular module containing MERs generated from ABMs without PAIR-C instruction. This comparison would allow us to draw a more solid conclusion on whether integrating PAIR-C features into ABM instruction could have added benefits in improving the visual contents represented in the original ABM, thereby fostering students' deep understanding of the inter-level causal relationships.

2) Future studies need to vary the pre-posttest items by including more open-ended questions, questions with multiple correct answers, and simulation-based question scenarios instead of only using multiple-choice questions.

The third is the use of prompt questions as a measure for deep understanding and misconceptions. The pilot study used two-tiered prompt questions embedded in the simulation module to elicit students' understanding of the inter-level causal relationships and their misconceptions. By coding the prompt responses, researchers found an underlying structure of students' misconceptions which aligns with the sequential features of the PAIR-C framework. This allowed researchers to assess students' understanding of the inter-level causal relationships during their learning process. Future studies should continue the use of prompt questions as a measure to capture and diagnose misconceptions revealed in students' causal framing. In examining students' learning process using prompt responses, researchers should also pay attention to whether students could make connections between knowing the four interaction features with reasoning the inter-level causal relationships.

Finally, is the selection of study participants. Given that the pilot study was conducted during the first wave of COVID, participants recruited were limited to a group of highly competent high school students attending college-prep summer programs. Therefore, they possessed a relatively higher prior knowledge of the content area before joining the experiment, leaving limited space for significant improvement. In future studies, the recruitment of participants should be more representative and include students from diverse backgrounds.

To summarize, Pilot Study 1 has scholarly implications for online instruction. The results highlight that applying the PAIR-C framework in designing well-structured online modules provided an alternative way of thinking and learning complex emergent concepts with agent-based models. To further extend this line of work, future works will modify more regular ABMs and create MERs to complement the use of interactive ABMs in explicitly teaching the seven PAIR-C features. Future works will also conduct a stand-alone simulation study to validate the effectiveness of the PAIR-C module compared to the Regular module in terms of improving a deep understanding of the inter-level causal relationships when learning emergent process concepts, and systematically addressing robust misconceptions.

Pilot Study 2

Introduction

While there were promising results for the third learning module implemented in Pilot Study 1, the PAIR-C project team wanted to improve the Process module and the Natural Selection module. For the Process module, the team focused on eliminating PAIR-C terms and replacing them with more explanations connected to everyday examples. For the Natural Selection module, the team again removed the terms and focused on connecting the learning of natural selection with parallel explanations made in the Process module.

Moreover, the team redesigned the assessment by excluding items that were non-discriminatory in pilot study 1 and then included different types of questions. As a result, there were ten multiple-choice questions with single correct answers which were identical to pilot study 1 (aka. MC questions); three multiple-choice questions with multiple correct answers (aka. MCC questions), and five open-ended questions (OE).

Table 4 illustrates the sequence of the study.

Table 4.

Sequence of Pilot Study 2

Sequence of Events	Experimental Group (n=18)
1	Natural Selection Pretest
2	Process module
3	Process test
4	Natural Selection Module
5	Natural Selection Post-test

Pilot study 2 was conducted virtually in partnership with a local high school classroom on Honors Science Research with a small group of students ranging in age from sophomore to senior (16 – 18yrs old). Some of the students in the classroom had taken pilot study 1 over the summer, which rendered only 18 students eligible for this study. The study provided students with 2-3 hours of instruction.

Findings

Paired samples t-test showed that there was no statistically significant difference between the pretest and the post-test in terms of the ten MC questions. However, students’ posttest scores were significantly higher on the combined MCC questions and the OE questions compared to the pretest. Besides, we identified fewer misconception types but a higher frequency of misconceptions in the posttest compared to the pretest. There were three main kinds of misconceptions shown in students’ OE responses in the pretest, which were *alignment, different sets, and cumulative summing*. In the posttest, most of the misconceptions were on *different sets* and *alignment*.

Because we did not have a control group in Pilot Study 2, we used the control group data in Pilot Study 1 to compare the results. Since both studies had 10 identical multiple-choice questions and 1 similar open-ended prompt question, we conducted ANCOVA using pre-test scores as covariates. We found no significant difference between the control group (Mean = 7.69, SD = 1.75) and the intervention group (Mean = 7.29, SD = 2.37) for their performance on the post-test 10 multiple-choice questions. Because we labeled item difficulty level based on the PAIR-C attributes, we later conducted ANCOVA on five hard items which were subsets of the 10 multiple-choice questions. Using pre-test hard question scores as covariates, no significant difference was found between the control group (Mean = 3.85 SD = 0.99) and the intervention group (Mean = 3.53, SD = 1.66).

For one similar open-ended prompt question, we coded students’ responses and identified misconceptions based on the PAIR-C framework. Results revealed that students from the control group had demonstrated five cases of *different set* misconceptions out of 14 responses (36%) while the intervention group had demonstrated six cases of *different set* misconceptions out of 18 responses (33%).

Students’ scores on the Process test and their scores on the Natural Selection pre-post tests were found to be strongly positively correlated: $r = .739$, $p = .001$ (for process and pretest correlations); $r = .769$, $p < .001$ (for process and posttest correlations). Multiple regression was run to examine how the post-test score could be explained by the pre-test score and the process test score. Although the two variables could explain a large variance for the post-test scores, $F(2, 14) = 11.59$, $p < .01$, R square = .623, only students’ process score added statistical significance to the model, $p = .034$ (See Table 5).

Table 5.

Multiple Regression Model for Pilot Study 2

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	7.35	2.98		2.47	0.03
Pretest Total Score	0.27	0.24	0.27	1.09	0.29
Process Total Score	0.51	0.22	0.57	2.34	.034

Dependent Variable: Posttest Total Score

Discussion

Although Pilot study 2 did not include a simulation module, it still provided important information and caveats for future studies to consider.

First, one reason for not finding statistically significant improvement and difference for the intervention group was because there was too much noise in the data collected. Because students received and completed the test without being monitored, the length of time for them to finish the tests were unreasonably longer or shorter after retrieving the time logging information from the Qualtrics survey. It was found that 7 students had excessively long lengths of time taking the pretest. For instance, two students had the test open for 5 days. For the posttest, one student took the exam in 50 seconds and one other student took the exam over 9.4 hours (the exam was left open overnight). Having excessively long or short test-taking students among a small sample was detrimental to the study for getting valid results.

To prevent this from happening in future online studies, it is critical to monitor students' test-taking process as well as their learning process. Most importantly, students who participate in this study should not be only driven by monetary incentives. Participants should be motivated to learn and take accountability for their learning and test-taking performance.

Second, it seemed that including diverse forms of test items (i.e., multiple-choice questions with multiple correct answers, and open-ended questions) allowed to capture more differences in terms of student deep understanding and misconceptions. Therefore, future research should include more items that were open-ended and require thoughtful considerations.

Third, when analyzing students' misconceptions, it did not make sense to simply count the frequency of misconceptions because one student could express one type of misconception repeatedly. Future research should focus more on characterizing different categories of misconceptions from the lens of PAIR-C and dedicate to comparing different patterns of misconceptions expressed before and after the experiment for each group.

Fourth, the strong and positive correlations between the Process assessment and the knowledge tests indicate that it might be possible to integrate the two concept learning modules (aka. combining the learning of the concept-general process module with the learning of the concept-specific module) into the simulation-based learning module. It might also be possible to assess students' understanding of the general process concepts in the same test as items for knowledge transfer. If this integrated learning module works to improve students learning of natural selection, it will fulfill the research goal of transforming complex system teaching into flexible and efficient online modules.

Conclusion

The two pilot studies described above both focused on teaching *natural selection* through the PAIR-C framework in an online environment. Looking across the two pilot studies, the lack of significant results on conditional differences could be due to two main reasons: 1) the age of participants being in the range of 14-18 showed a lack of self-regulation during the self-directed online learning process, and a tendency to join the experiment for a monetary incentive; 2) the intervention was brief and did not provide enough time or scaffolding to help students digest information and transit from one learning module to another. Given that it is still difficult for students to reach a deep understanding and alleviate common misconceptions of natural selection, this dissertation study continues this line of work by conducting another online study to compare the effect of using the PAIR module versus the Regular module in facilitating a deep understanding of natural selection.

CHAPTER 4

METHODOLOGY

Study Design and Setting

The study adopted a pretest-posttest randomized block design (RBD). The blocking factor was the participants' scores on the pretest questions. To be specific, I ranked students according to their scores on the pretest True or False (T/F) questions and divided them based on ranks into two equal-sized groups. Subjects in the two groups, known as blocks, were less diverse in their understanding of natural selection than those in the condition of an unblocked design. Once the blocks were formed, I randomly assigned an equal number of members from each block to each of the two conditions, and the experiment proceeded. By adding the blocking factor, error variance was reduced, which increased the internal validity of the experiment. RBD assured the baseline equivalence of subjects among the two treatment groups so that the detected group difference could be more confidently attributed to treatment effects (p.228-231, Wickens, & Keppel, 2004).

Context

The context of the present study was a Technology in Action project offered in a technology literacy online course as an alternative assignment project at a US Southwestern University during the summer session of 2022. This course spanned over seven weeks and was in asynchronous mode in which all learning materials were prescribed and prepared by five instructors and one instructional designer. Each instructor was responsible for providing feedback to a separate group of students. There were a total of 108 students enrolled in the course.

During the Fall of 2021, I taught this technology literacy online course myself. In addition, I participated in the course design meetings with the lead instructor and the instructional designer to brainstorm ideas for increasing students' engagement by re-designing the course contents. Since I was researching on agent-based simulations, I proposed the idea of introducing ABM to this course and providing an opportunity for online students to take part in an active research study that aims to test the effectiveness of simulation technology in learning crosscutting concepts. This idea was well accepted by the whole instructional team and was later implemented during the summer session.

Participants

Although there were about 80 students joined the study, only 50 of them signed the consent forms and completed all instructional activities, pre-posttests, and surveys. Among the 50 participants, 26 of them were in the Control group and 24 were in the Intervention group. A short demographic survey was used to track their self-identified gender, age, racial ethnicity, educational background, major of study, biology learning background, and prior experience with simulation. The gender distribution was 76% female students, 22% male students, and 2% non-binary. All participants were over the age of 18 years old. The average age was 27.4 years with a standard deviation of 8.91. Participants reported their ethnicity with the following proportions: 68% as White, 14% as Hispanic/Latino, 6% as Asian or Asian American, 6% as Black or African American, 2% as Native American, and 4% as Other. They were juniors (52%), sophomores (22%), seniors (16%), and other levels (10%), majored in Education (n = 29, 58%), Social

Studies (n = 13, 26%), and Arts & Humanities (n = 8, 16%). A total of 86% of participants reported having taken 1-2 biology classes at high school, 12% reported having taken 3 or more biology classes at high school, and only 2% reported having taken no biology class at high school. Only two participants reported having heard about and used NetLogo simulation before the study.

ABM Tool – NetLogo Models

NetLogo (Wilensky, 1999) was the selected ABM tool in the present study. As one of the offspring software of LOGO, NetLogo possesses all essential features of ABM tools and can simulate interactions among agents, which give rise to emergent phenomena over time (Wilensky & Resnick, 1999). NetLogo models refer to ABMs created using the NetLogo multi-agent-based programming environment, which often includes thousands of agents moving simultaneously and independently in a system. By setting the behavioral rules of these agents, students will have the opportunity to study the features and potential mechanisms of an emergent phenomenon. In addition, NetLogo is a relatively reliable tool. It has been developed for more than twenty years and has been used in many educational research settings (Blikstein & Wilensky, 2009; Stroup & Wilensky, 2018; Wilensky & Rand, 2015; Wilensky, 1999). Finally, NetLogo is freely available via download and is compatible with the operating systems on both Macs and PCs. It can also be operated in a web-based browser environment using NetLogo Web. Thus, it is an economical and practical tool for implementing this online study.

Target Learning Content – Natural Selection

Natural selection (NS) was chosen as the target learning content in the present study for three reasons. First, NS is one of the fundamental concepts of modern biology and addresses an essential phenomenon of living things—biological evolution (National Academy of Sciences, 1998).

Second, NS is a content area where new learning methods are still needed. Although many studies have examined a variety of tools (mostly simulation tools) to scaffold students' understanding of natural selection (Abraham et al., 2008; Bray et al., 2008; Fiedler et al., 2018; Horwitz et al., 2013; Pope et al., 2017), students at all levels - not only naïve learners but even advanced learners who have received postsecondary biological education and training - often experience difficulties in coming to a deep understanding of NS in part because they hold robust misconceptions or alternative Lamarckian conceptions (Bishop & Anderson, 1990; Gregory, 2009; Nehm & Reilly, 2007). According to Chi et al. (2012a), students, in general, have little knowledge about emergent-causal schema and are intuitively familiar with the direct-causal schema. Therefore, it is extremely hard for them to correctly identify NS as an emergent process and use an emergent-causal schema to explain the causal mechanisms of NS.

Third, NS is a good content area to infuse ideas of emergence using ABMs. NS is ontologically an emergent phenomenon in which population changes over time emerging from the behaviors of individual living organisms and the interactions between these organisms and their environment. In addition, it is also a level-based phenomenon. One can think of variation and selection as occurring at levels of the gene, the individual, or the species. The levels-based ontology of NS makes for a good match with ABMs that foreground the notion of levels and the importance of understanding how interactions at one level might lead to changes at another level (aka. inter-level causal relationships).

Equipment

During the study, students worked individually using their laptops with stable internet connections. Each student accessed the study materials from their course Canvas shell. When students encountered technical challenges, they could ask for assistance from the course Slack channel or directly email or call the researcher. Detailed use of different software to collect data will be described in the section below.

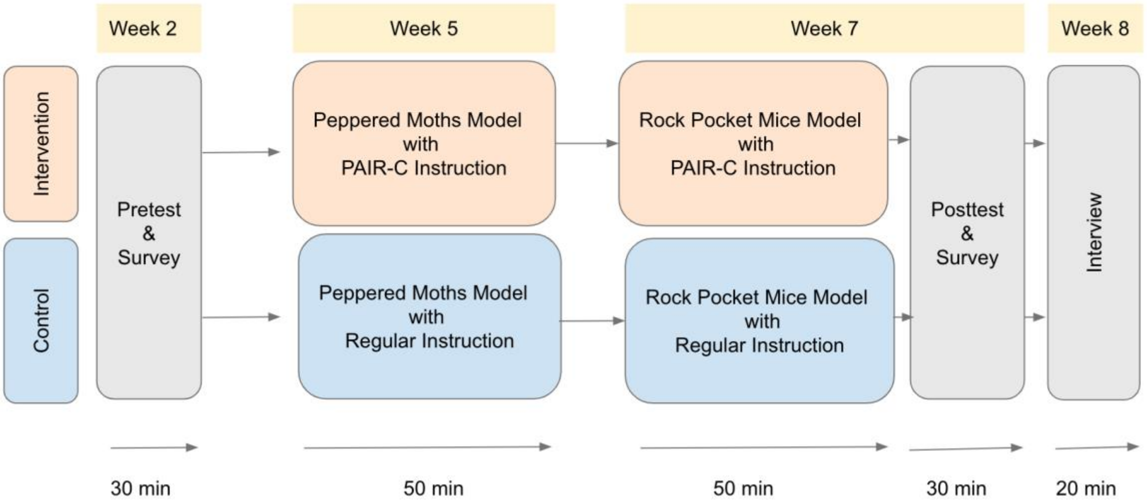
Data Collection

Procedures

The data collection took place over 8 weeks during the summer of 2022. Figure 4 showed the overall study process and procedure.

Figure 4.

The Overall Study Process and Procedure



Before the study started, all students enrolled in this online course watched a video recording in which the lead instructor introduced the Technology in Action (TIA) project and encouraged students to choose this project in replacement of the regular course assignments. Interested students then entered the first TIA module on Canvas which contained the study overview and the consent form. Students who decided to participate signed the consent form.

During week 2, students were given a pretest followed by a background questionnaire. After collecting the pretest results, students were assigned into two groups using the RBD approach. A roster was created to inform the instructor team and the instructional designer about pretest completion and group assignment results. Based on the roster, the instructional designer created two conditions under the TIA project Canvas page and assigned each student to their corresponding condition. As a result, students assigned to condition 1 would only have access to the PAIR-C modules whereas students in condition 2 would only receive the Regular modules.

During week 5, participants in each condition were given access to the corresponding learning module. At the beginning of the module, participants were asked if they were willing to record their experiment using Zoom. Guidance on how to record and share the recordings was provided to those who agreed to record. Participants in both groups also received tutorials on how to navigate the NetLogo Web to run the Peppered Moths ABM and the Qualtrics-supported instructional page.

During week 7, participants in each condition were given the corresponding second module in the context of Rock Pocket Mice. They were also asked to record their experiment and were provided with instructions. The post-test and the post-project questionnaire were given immediately after the module ends. However, there was a 10-minute break reminder before the start of the posttest. In the post-project questionnaire, students were asked if they were willing to participate in a 20-minute interview after the study. Those who agreed to participate in the post-study interview were prompted to select a time slot from all available times within a 2-week window. The researcher sent out an email notification and a Zoom invitation link to participants who signed up for the interview. All interviews were recorded upon participants' permission.

At the end of the experiment, participants who finished all the learning materials received a certificate of completion issued by the Learning and Cognition Lab. All instructional materials (e.g., questionnaires, pre-posttests, simulation prompts) were delivered using Qualtrics and embedded into the Canvas TIA project page.

Data Sources

To answer the research questions, I focused on collecting data on students' learning outcomes and learning processes. The main data sources used in the study included 1) pre-posttest, 2) prompt question responses in the learning module, 3) students' recorded videos of their experiments, 4) pre- and post-study questionnaires, and 5) post-study interviews.

Instruments

Pretest and Posttest.

Students' understanding of Natural Selection and other emergent phenomena was assessed by the pretest and the posttest instrument. Each test consisted of two sections and required students to take 20-30 minutes to complete. Section A of the test contained three sets of True or False (T/F) questions on natural selection. Each set was accompanied by five possible statements. There were a total of 15 T/F items designed in the format of two-tiered questions. The first tier simply asked participants to decide "True" or "False" for each statement. The second tier was open-ended questions which prime participants to provide explanations for the statements they identified as false. The two-tier questioning strategy was used to diagnose students' misconceptions in their reasoning. Section B of the test contained four open-ended (OE) questions with the first two questions assessing students' understanding of *natural selection* in different contexts (to be referred to as context transfer items). The last two OE questions were used to assess students' ability to transfer their understanding of inter-level causal relationships from learning *natural selection* to other emergent phenomena (to be referred to as far transfer items).

After the completion of the two modules, the same instrument with different orders was administered as a posttest to measure any change in students' understanding. The only difference between the pretest and posttest is the addition of two sub-questions related to the two far-transfer items. These newly added far transfer items provided students with ABM simulation videos and asked students to explain how the simulation can help them understand the process of geese flying into a V-shape, and the process of fake news spreading. The entire pre-posttest instruments are available in Appendix A.

Validity and Reliability of the Pretest and Posttest.

The pre-posttest instrument used in this study measures one primary construct: understanding the inter-level causal relationships (ICR) of natural selection. According to the PAIR-C framework, understanding ICR for emergent process means understanding how the pattern is produced from a converging process involving all the agents' interactions at each unit of time through collective summing causal mechanisms (see Feature 5 and 6, Table 1). Moreover, the understanding is often manifested in two implication statements students often made when explaining emergent causality: no controlling agent(s) responsible for the pattern; no alignment between the behavior of the agents and the behavior of the pattern (Features 7a and 7b, Table 1). Therefore, assessing students' understanding of the ICR for natural selection involves understanding the inter-level PAIR-C features and the implications features in the context of natural selection. Based on this principle, items were selected from previous PAIR-C studies and were revised to measure the primary construct.

One source of evidence for assessing validity and reliability came from hiring a content expert to review the test and provide feedback on the test content and construct. The content expert was a high-school biology teacher who previously taught the PAIR-C framework in his natural selection lessons. He carefully examined whether the test could assess students' understanding of ICR for natural selection and suggested changes to the question content, wording, format, and distribution of questions focusing on different PAIR-C features. His suggestions helped resolve issues such as construct underrepresentation or construct irrelevance. Appendix B provides a cross-tabulation of the construct for each item in the finalized assessment. Appendix C provides an overview of each question which gives information on the difficulty level, question source, historical use of the item in previous PAIR-C studies, and the format of each item.

Since subjective judgment was involved in the test scoring process, inter-rater reliability was established between raters that code and score for students' open-ended (OE) responses (p.44, Standards 2.7, AERA, 2018). Kappa was reported as evidence for inter-rater reliability.

Prompt Questions.

In this study, participants in both conditions were provided with two learning modules along with two different simulation scenarios. In each module, they responded to 20 prompt questions. These prompts were designed based on the ICAP framework (Chi & Wylie, 2014; Chi et al., 2018) and could be categorized into two types: active prompts and constructive prompts. Active prompts asked students to manipulate simulation parameters, select answers, arrange statements, observe simulations, and describe patterns. Constructive prompts were often open-ended and asked for an explanation, a prediction, or a reflection. By adopting the ICAP framework, the researcher could measure students' cognitive engagement level by analyzing whether students responded to a prompt in a passive, active, or constructive manner. Moreover, some prompt questions were also content questions that could help the researcher assess students' understanding as well as capture and diagnose misconceptions revealed in students' causal framing.

The prompt questions for each module were identical across two conditions to ensure that students in both the PAIR-C intervention condition and the Regular control condition were given equal opportunity to be actively engaged with the simulation and to respond constructively to open-ended prompts that promote deeper learning and understanding (Chi et al., 1994). Appendix D shows all the prompt questions used in the two modules.

Questionnaires.

Participants completed a short background questionnaire before the experiment and after the pretest. The background questionnaire included the following demographic information: gender, age, ethnicity, mother tongue, educational level, major, number of biology classes taken in high school, and prior experience with the NetLogo simulation.

Participants completed another short questionnaire after finishing the posttest. The post-study questionnaire included the following elements: (i) Three five-point Likert scale items that asked participants to rate the degree of whether they thought the module and the instruction were engaging, helpful, or clear. Items ranged from (1) Not Engaging/Not Helpful/Not Clear to (5) Very Engaging/Very Helpful/Very Clear. (ii) Three open-ended questions asked students to describe the confusing part of the module, compare the NetLogo simulation with other simulations they've played before, and give general comments about the Technology in Action project. (iii) Self-reported time duration for completing each module. (iv) Willingness to participate in a 20-minute post-project interview. Both questionnaires can be found in Appendix E.

Post-Study Interview Protocol.

The interview protocol contained four parts. Part One asked questions based on students' responses to the questionnaires. For instance, if students majoring in education, the interviewer would ask them what subject they teach, and how they intend to use simulations in their teaching and ask them to describe a teaching scenario using simulations. Part Two asked questions that could help the researcher better understand students' learning processes and learning outcomes. For example, students were asked to reflect on their changes in understanding the concept of natural selection. Part Three focused on questions related to the use of simulations which would inform future design iterations of the PAIR-C module. When interviewing participants from the PAIR-C condition, one question directly asked their thoughts on using "links" to visualize agent interactions. Part Four focused on questions related to learning transfer and real-world applications. It also asked students to indicate their self-efficacy in using simulations to understand or explain complex concepts. By the end of the interview, students were also allowed to raise questions about the project. The post-project interview protocol can be found in Appendix F.

Data Analysis

I have combined quantitative and qualitative methods in the data analysis. To construct the coding scheme, the students' answers to open-ended questions were analyzed through a combination of theory-derived and data-derived analysis (Chi, 1997). The coding scheme was then used to score and analyze the whole corpus of data quantitatively.

Coding and Scoring the Pretest and Posttest

For the True/False selection part of the question, if students correctly determined whether the statement was True or False, they were given a score of 1, otherwise a 0 point. For the open-ended part of the T/F items, we created scoring rubrics for each false item (on a scale of 0, 0.5, and 1). Note that there were 10 false items out of the 15 T/F items. Similarly, we also created scoring rubrics for each transfer item to assess students' transfer performance on a scale of 0 to 5. For each open-ended item, more than 50% of student responses were scored by two researchers for inter-rater reliability over three rounds (Kappa for the open-ended part of the T/F questions was an average of 0.843, $p < 0.001$, Kappa for OE transfer questions was 0.925, $p < 0.001$, indicating strong scorer agreement). Appendix G shows the scoring rubrics for all the open-ended test questions, along with sample responses from students.

In addition to the scoring rubrics that were used for rating students' OE responses. A two-dimensional coding scheme was developed to identify and categorize students' misconceptions that appeared in their OE responses. The first dimension of the coding scheme consisted of nine misconceptions: MC1: intra-generational change, MC2: use and disuse, MC3: directed variation, MC4: teleological conceptions, MC5: anthropomorphism, MC6: essentialist conceptions, MC7: transformational views, MC8: events vs. processes, MC9: absolutes vs. probabilities. The second dimension of the coding scheme was based on the PAIR-C features under sequential processes: Feature 1: distinct sets of interactions, Feature 2: restricted interactions, Feature 3: serial order; Feature 4: dependent interactions, Feature 5: incremental change, Feature 6: cumulative summing, Feature 7a: controlling agent, Feature 7b: align. This first dimension of the rubric was designed based on an extensive review of natural selection misconception literature (Abraham et al., 2009; Gregory, 2009; Nehm & Reilly, 2007; Shtulman, 2006) and the second dimension of the rubric was designed based on the PAIR-C framework. Codebooks are available in Appendix H and Appendix I.

When coding for misconceptions, students' OE responses were reviewed and coded by two researchers. If a response contained a misconception, then the coder would mark "1" under the corresponding misconception category. If a response contained multiple misconceptions, then the coder would first determine whether these misconceptions were under different categories and only mark "1" for misconceptions of different categories. After assigning "1" for the existing misconceptions, the coder would fill in “0” for all cases that have no misconception. A similar strategy was used to code the presence of PAIR-C sequential features in each response. There were 20% of the responses coded by two coders. All disagreements were discussed and resolved in three meetings and asynchronously through Google Sheets. For unsure codes, both coders agree to generate two new categories to capture “undecided misconceptions” and “undecided PAIR-C features”.

Coding and Scoring the Prompt Questions

Students in both the intervention group and the control group received a total of 40 identical prompt questions embedded throughout the learning modules. All the prompt questions were involved in the coding process. Unlike the pilot study, this study did not code for content understanding and misconceptions. Instead, this study adopted the ICAP framework (Chi and Wylie, 2014) and created a coding scheme for assessing students’ level of engagement during the simulation learning process. In other words, how much effort students put into answering each prompt would demonstrate their level of cognitive engagement.

Note that these prompt questions were created with the ICAP level in mind. For example, active prompts asked students to describe their observations whereas constructive prompts asked students to explain, predict, or reflect. Hence, although each prompt question had its coding rubric, there was a general principle that could apply to coding most responses. For the active prompt questions, if students not only described but also extended their descriptions using prior knowledge, then their responses would receive a mark of “3” for constructive engagement. If they just described some details in their responses, they would receive a mark of “2” for active participation. If they merely or very roughly described the observation in their response, then they would receive a mark of “1” for passive participation. For the constructive prompt questions, if students reflected on the instructional text and pointed out what was missing in their response, they would receive a mark of “3” for constructive engagement. If they paraphrased part of the instructional text, then they would receive a mark of “2” for active engagement. If they just copied the instructional text, then they would receive a mark of “1” for passive participation. As participants in this study were online students who complete each activity and the project individually through their laptops, *interactive engagement* was not evaluated in this study.

Following this coding procedure, each student’s received a total score for engagement. Moreover, individual student engagement level was later calculated based on the ICAP percentage. The percentage of Constructive engagement (constructive%) was calculated by the number of constructive codes divided by the number of all codes. The percentage of Active engagement (active%) was calculated by the number of active codes divided by the total number of codes. Similarly, the percentage of passive engagement (passive%) was calculated by the number of passive codes divided by the total number of codes. For the student engagement codebook, see Appendix J.

Quantitative Approaches

Descriptive and inferential statistics were used to examine students’ performance on the pre-posttest. A one-way ANCOVA using pretest scores as the covariates was later conducted to determine whether there was a conditional difference between the intervention group and the control group on the post-test scores. Results for the pre-posttest were reported in Chapter 6.

Epistemic Network Analysis (ENA) was applied to model the structure of connections in the coded misconception data (Shaffer, 2017; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017). The ENA model included the following codes: MC1, MC2, MC3, MC4, MC5, MC6, MC7, MC8, MC9, F1, F2, F3, F4, F5, F6, F7a, and F7b (See the codebook in Appendix H and I). In the model, I aggregated networks using a binary summation in which the networks for a given line reflect the presence or absence of the co-occurrence of each pair of codes. The resulting networks are aggregated for all lines for each unit of analysis in the model. The results of the ENA analysis were reported in Chapter 7.

In addition to t-tests, correlation analyses, and multiple regression analyses were conducted when analyzing student engagement levels during the learning processes and their associations with learning outcomes. A video analytics tool, V-note, was also used to quantify the video data and provide triangulation for interpreting student learning performance in Chapter 8.

CHAPTER 5

DESIGN OF LEARNING MODULES

In this chapter, I describe the design and development of the PAIR-C module and the Regular module used in the study. PAIR-C module incorporated the seven PAIR-C features into the instruction and was used in the Intervention group. Regular module applied instruction used in common ABM integration practices and was used in the Control group. Much of the work on these two modules was inspired by the BEAGLE (Wilensky & Novak, 2010; Wagh & Wilensky, 2013; 2018), and the CT-STEM projects (Kelter et al., 2021; Peel et al., 2019; Wilensky, 2018). Specifically, this study adopted two models and the corresponding curricula from these two projects: a) Peppered Moths - a model based on the famous example of the change of the coloration of moths in response to pollution (from BEAGLE); b) Rock Pocket Mice - a model simulates fur coat color changes in the population of rock pocket mice due to predation (from CT-STEM).

Both modules were about the same length and used the same amount of multiple external representations (MERs) at the same location including 2 interactive agent-based models, instructional texts of about 8,500 words, 7 images, 8 animations, 20 screenshots, and 10 videos. A total of 40 identical prompt questions were used to ensure students have equal opportunities to be actively and constructively engaged in their learning processes. In general, students in both groups could interact with the module by reading the instructional texts, observing simulation videos and images, exploring the model by adjusting different parameters, testing hypotheses, answering prompt questions, and generating new ideas and inferences beyond manipulating the simulation. Table 6 presents a detailed comparison between the instructional steps and contents of the Regular module and the PAIR-C module.

Table 6.

Comparison of the Two Conditions

Steps	**Regular Module (Control group)**	**PAIR-C Module (Intervention group)**
Week 2	The Peppered Moths Simulation Scenario	
Activity 1 Observe the simulation and make predictions. **(5 mins)**	• direct instruction on the simulation interface • 1 video tutorial • 2 prompts	• direct instruction on the simulation interface • 1 video tutorial • 2 prompts
Activity 2 Experiment with the simulation and learn about the influence of selection pressures on the changes in variation distributions. **(15 mins)**	• direct instruction on two Darwinian Principles (DPs) and the function of the Selection slider with examples elaborating how different selection pressure determines the survivability of the moth species • 1 simulation video without visual cues • 6 prompts	• direct instruction on the Pattern, Agents, Interactions, and Relations dimension with examples that emphasize all moths still have the same four interaction features despite selection pressure • 1 simulation video with blue links showing interactions • 6 prompts
Activity 3 Use the simulation to examine the role of environmental factors on the pattern of variation changes. **(15 mins)**	• direct instruction on how to avoid common misconceptions while explaining why the changes happen • 2 simulation videos without links, 2 ABM screenshots showing the initial pattern and the final pattern respectively • 7 prompts	• direct instruction on the "converging change" feature (under the Causality dimension) to explain why the changes happen • 2 simulation videos with links, multiple ABM screenshots showing the dynamic changing pattern across time • 7 prompts
Activity 4 Understand how mutation might	• direct instruction on mutations, the effect of mutations, and using DPs to	• direct instruction on using the "collective summing", "no-direct effect", and "not

result in changes in variation distribution. **(15 mins)**	explain how the changes happen at different mutation levels. • 1 simulation video without links • 5 prompts	always align" features to explain the causal mechanism of the changes • 1 simulation video with links • 5 prompts
Week 5	The Rock Pocket Mice Simulation Scenario	
Activity 5 Explore the simulation and check predictions **(10 mins)**	• direct instruction on the simulation interface • 2 video tutorials • 5 prompts	• direct instruction on the simulation interface • 2 video tutorials • 5 prompts
Activity 6 Experiment with the simulation and learn how genetic mutation might result in pattern changes. **(10 mins)**	• direct instruction on mutation and the use of two DPs to describe the natural selection process. • 2 simulation videos without visual cues • 6 prompts	• direct instruction on mutation and the four interaction features shown • 2 simulation videos with blue links showing interactions • 6 prompts
Activity 7 Examine how environmental change places pressure on the mice population, thereby influencing the pattern change. **(15 mins)**	• direct instruction on three other DPs, and how to avoid common misconceptions while explaining why the changes happen • 1 simulation video without visual cues to show interactions • 8 prompts with one asking to re-organize statements to elaborate the process of natural selection	• direct instruction on all the inter-level features (under the Causality dimension) to explain why the changes happen • 1 simulation video with red and blue links showing two types of interactions • 8 prompts with one asking to re-organize statements to elaborate the process of natural selection
Activity 8 Review key ideas to understand and explain the process of natural selection. **(15 mins)**	• review all the five Darwinian Principles • 1 prompt asking to select and explain three misconceptions from a list of six	• review all the PAIR-C features • 1 prompt asking to select and explain three misconceptions from a list of six

As shown in Table 6, the two conditions used different instructional materials. Specifically, the control condition provided students with direct instruction on the five Darwinian principles (individual variation, genetic determination, adaptation, reproductive advantage, accumulation of changes over time), conventional definitions, and common misconceptions but withhold information on the emergent properties shown in the simulations. Moreover, MERs used in the Regular Module were directly generated from ABMs with no modification and did not contain external visual cues to represent different types of agent-level interactions. In contrast, the intervention condition provided direct instruction on all the PAIR-C features and how to apply these features to explain the emergent properties of natural selection caused by mechanistic details of agent interactions (such as inheritance, predation, and reproduction). Most of the MERs used in the PAIR-C Module were modified and provided with visual cues (e.g., blue links represent mating interactions, and red links represent predation relationships).

In terms of similarities, both conditions initially provided direct instructions on DPs or PAIR-C features out of the context of Peppered Moth and Pocket Mice scenarios. As students spent more time interacting with the ABM simulations, instructions on DPs or PAIR-C features were later situated in contexts. Both conditions used the same number of identical prompts at the same location which allowed researchers to collect data from students' responses and formatively assess students' understanding as well as cognitive engagement. An equal amount of time was reserved for each condition to complete the same number of activities and instructional steps. Moreover, the activity designed under each simulation scenario was bounded by the interaction between students and different aspects of the simulation as well as different instructional goals. For instance, in Activity

1, students only observed the simulation without adjusting any parameters with the goal of familiarizing themselves with the simulation environment. In Activity 2, students were guided to adjust the *selection slider* of the simulation with the goal of understanding the influence of selection pressure. In Activity 3, students were guided to adjust the *pollution settings* of the simulation with the goal of examining environmental factors. In Activity 4, students were guided to adjust the *mutation slider* of the simulation with the goal of understanding the effect of mutations.

The PAIR-C Module

Before integrating the PAIR-C features into designing the modules for teaching the concept of natural selection, five issues were first identified from the PAIR-C perspective.

The first issue is that the **P**atterns in regular ABMs did not clearly represent dynamic changes across multiple generations. If one uses the example of peppered moths getting darker over generations, typically the simulation changes fast and points to the end state of moths having darker-colored skin.

The second issue is that the regular ABMs did not differentiate external **A**gents from internal agents. In the process of moths getting darker, for example, interactions of moths (internal agents) can occur with predators such as birds, but the birds are “external” agents that do not participate in the *Pattern* of darkening moths.

The third issue is that agents’ **I**nteractions (e.g. mating interaction; chasing interaction) were not visible in the regular ABMs and only the agents’ actions (e.g. agents randomly move) were shown. Because natural selection is a complicated concept that involves different types of interactions (or a set of interactions) among the agents (e.g., moths not only reproduce with each other but are also being preyed upon by birds, etc), it is critical to visualize different types of interactions to promote understanding.

The fourth issue is that **R**elations among the interactions are not presented and explicitly stated in the regular ABMs. For example, students are less likely to infer that a light-colored moth could continue to interact with other light-colored moths (because interactions are random) even the moths are getting darker.

The fifth issue is that the **C**ausal relationships between the agents-pattern (how the pattern converged and how the pattern was produced) were not explicitly explained in regular ABMs. To be specific, after industrial pollution, more light-colored moths are getting spotted and eaten by birds. It does not mean that all the light-colored moths will die and all the dark-colored moths will survive, resulting in the pattern of moths getting darker. Instead, dark-colored moths can still reproduce with light-colored moths, resulting in another generation of not necessarily dark-colored moths. So, the final pattern of moths getting darker is not derived from a simple conditional relationship, such as "able to escape from predators, therefore will survive and reproduce."

In the following section, I will elaborate on how the PAIR-C features were integrated into the design of the PAIR-C module while addressing the five issues mentioned above.

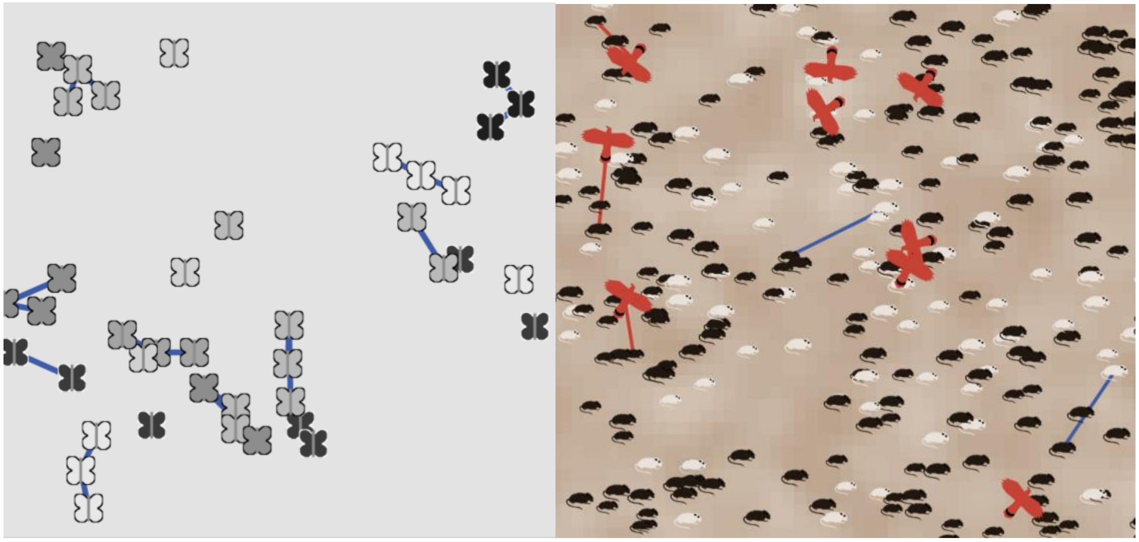
Integrating the Four PAIR-C Interaction Features

Through integrating the PAIR-C interaction features, the third and fourth issues were addressed. To start with, the interactions between agents were first made visible by adding rules such as "directed-link-breed". These links were added in both the Peppered

Moths and the Rock Pocket Mice models. As shown in Figure 5, mating interactions in both models were represented using blue links. Predator-prey interactions were represented using red links in the mice model. The “links” were identical in terms of thickness and were non-directional. Notice that students in the PAIR-C intervention group did not interact with the modified models directly. They interacted with the same regular models as the control group. However, the instructional materials (such as the videos, images, screenshots, and instructional texts) used in the PAIR-C group were created based on the modified models.

Figure 5.

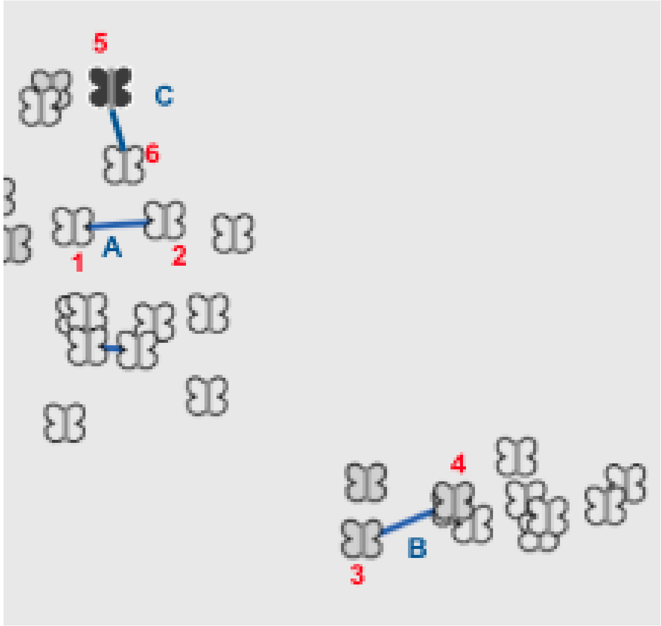
The View of the Two Modified ABMs (with links shown)



Therefore, the visible “links” in the MERs enabled the instruction on the four interaction features. For example, one image from the Peppered Moths PAIR-C module presented three pairs of moth interactions with links to illustrate the interaction features (See Figure 6). In the image, the researcher marked the three pairs of mating interactions as A, B, and C, and also assigned numbers (1-6) to the six moths.

Figure 6.

An Image Showing PAIR-C Interaction Features



The four features were presented and illustrated with examples below:

- Feature 1: Same/uniform set of interactions
    - Both dark-colored moths and light-colored moths could get eaten by birds, and all moths who survive the birds can mate and reproduce offspring.
- Feature 2: Random or unrestricted interactions
    - Moths who survived can mate with any other moths despite the color. For example, light-colored moths can not only mate with light-colored ones but also with dark-colored ones (see the mating interactions: A and C)
- Feature 3: Simultaneous interactions
    - Moths' mating interactions co-occur at the same time. For example, Moth 1 & Moth 2, Moth 3 & Moth 4, and Moth 5 & Moth 6 can mate at the same time.

- Feature 4: Independent interactions
    - A Moth in one location mating with another moth has no effect on any other moths' mating interactions. For example, mating interaction C does not depend on other mating interactions.

After introducing the four interaction features, students received a prompt asking them to run the simulation and explain how the distribution change of color variations was influenced by selection pressures. A possible answer was given to students after they entered their responses. In the solution, the four interaction features were mentioned again because all moths still had the same four interaction features despite changes in the selection pressure. Upon revealing the answer to students, another prompt primed them to reflect on what information on the answer page was most important for understanding. This reflection prompt is meant to help students notice the importance of understanding the four features in explaining the emergent process.

The four interaction features were explicitly taught to students in the PAIR-C intervention group multiple times. In most cases, instructional texts about these four features were provided simultaneously or immediately after students interacted with the link visible MERs (e.g., videos, screenshots, images, etc.). For instance, from the MERs, it was obvious for learners to see that an agent could interact with any other agents randomly. Given that students often thought a subgroup of fit moths only reproduced with other fit ones (e.g., the light-colored moths only reproduced with light-colored ones), teaching the perceivable features that mating interactions occurred randomly, simultaneously, and independently could help students avoid treating moth interactions to be restricted within subgroups. Since Chi (2012b) suggested that the four perceivable interaction features were foundational for understanding inter-level features, instructions about the four features were given before mentioning the inter-level features and appeared three times in the first Moth simulation scenario and one time in the second Mice scenario.

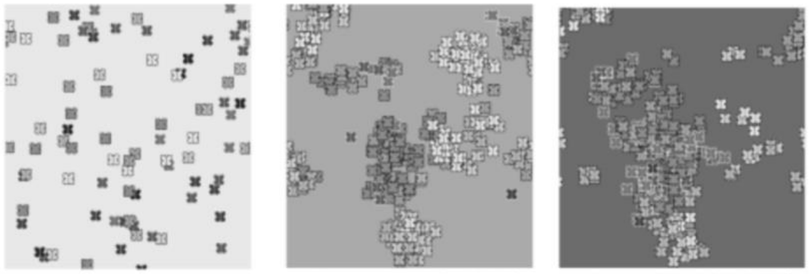
Integrating the Two PAIR-C Inter-level Features

Integrating the PAIR-C inter-level features were conducive to addressing the first and fifth issue. In addition to using the MERs to illustrate the four perceivable interaction features, they were also used to show the dynamic changing pattern as well as instruct the PAIR-C inter-level features. Specifically, students in the PAIR-C intervention group were first instructed to run and pause the Moth simulation at six different time points (tick 0, tick 25, tick 50, tick 75, tick 100, tick 125). Then, they were asked to take and upload the screenshot for the simulation at tick 0 and tick 125 respectively. Next, they were prompted to explain how the color variations at Tick 125 were different from Tick 0. A possible answer with six ABM screenshots was given to students which showed moth color change over six generations (Figure 7).

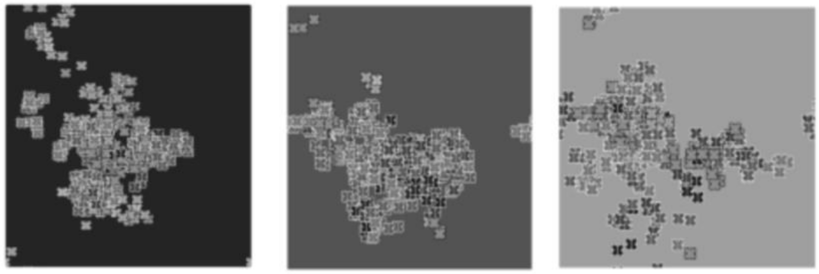
In addition to observing the multiple screenshots captured from Generation 1 to Generation 6, instructional texts pointed out that the color variations of moths were almost evenly distributed at tick 0 whereas dark-colored and medium-colored moths were most seen at tick 125. This indicated that the initial pattern at Generation 1(tick 0) was not part of and had no resemblance to the final pattern at Generation 6 (tick 125).

Figure 7.

ABM Screenshots Showing the Converging Pattern over Six Generations



Generation-1 (Tick 0) Generation-2 (Tick 25) Generation-3 (Tick 50)



Generation-4 (Tick 75) Generation-5 (Tick 100) Generation-6 (Tick 125)

Based on this observation, a follow-up prompt asked students to explain why Tick 0 looked different from Tick 125. The possible answer for this prompt question hereby introduced the term "converging change" (Feature 5, Table 1) to qualitatively describe the changing pattern and then provided a more elaborative explanation to students. The instructional text stated "To explain why the pattern at Tick 125 looks different from that at Tick 0 (why the changes happen), we need to consider interactions of all the agents (i.e., moths mating interactions; moth-bird interactions) at each generation. Specifically, at any generation, the changing pattern reflects all the moths' interactions, such as dark-colored moths mating with any random other moths and giving birth to moths with different colors. The birds can eat any moth, no matter the color. We also need to understand that all interactions approximate the final pattern over time. One cannot predict the exact number of moths being survived and reproduced in each generation, but it is possible to create approximate predictions." The instructional text continued the elaboration "We know that dark-colored moths are less likely to be spotted and eaten by birds in the soot-filled environment and tend to survive and reproduce more offspring than light-colored moths. Therefore, one can predict that the proportion of moths that are dark are likely to increase over time relative to those that are light-colored."

In the first Moth scenario, the PAIR-C module also provided explicit instruction on the "collective summing" causal mechanism (Feature 6, Table 1). Specifically, students were given a link-visible ABM video that recorded moths' color change under three different mutation levels (from low to high) with selection pressure at 50% and pollution level at 90%. Then, they were asked to explain how the pattern changed over time. From the simulation video, students could observe that when the chance of mutation increased from low to high, there was a trend of increasing light-colored moths. In the possible answer, instructional texts introduced the term "collective summing" to quantitatively describe the causal relationships between agents and the resulting patterns. The texts also stated that "To further explain how the changes happen, you need to understand that the pattern of light-colored moths increasing is the collective outcome of all the moths interacting at the local level." Followed by this statement, a more comprehensive answer was provided and explained "light-colored moths becoming common is the net effect of adding positive and negative numbers of light moths considering all the interactions within that generation and then comparing across generations (e.g., adding positive interactions: light moths reproduce and give birth to a light moth; adding negative interactions: light moths fail to escape from birds and die). However, the pollution level remains to be high at 90%, and the overall pattern still shows that light-colored moths are less common in the moth population compared to other types of moths.”

Integrating the PAIR-C Implication Features

To illustrate the implication feature of “all agents are responsible for the pattern” (Feature 7a, Table 1), students were first guided to adjust the simulation parameters (Turn off the “cycle-pollution” switch and set the pollution level to be at 90%; Make sure the “selection” slider is set to 50, and there are at least 100 moths in the world;
Set the “mutation” slider to 50) and record the percentage of dark moths at different ticks shown in the monitor boxes. After this activity, students were asked to decide True or False for the following statement: “If a newborn moth has a mutation, its color will always be darker than its parents” and then explain their responses in 2-3 sentences.
When students were collecting the data from observing the simulation, they noticed that the percentage of dark-colored moths was not always increasing although the pattern showed that moths were getting darker over time. This observation was critical in helping students respond to the True or False question.

After their responses, students received the following answer in the text “The statement is false because although darker-colored moths are more common over time, they **cannot control or directly affect** the body color of their offspring. Mutations occur randomly and light-colored moths can still reproduce if they live to adulthood, so it is possible that there are still light-colored moths being born in the process. Moths cannot pick or control their phenotypes to be advantageous.” This answer was followed by instructional texts which highlighted that “No direct effect” was one of the inter-level implication features in the PAIR-C framework. The instructional texts also pointed out that “To understand this inter-level implication feature, you need to also know:

1. One cannot attribute a single moth or a group of moths, such as dark-colored moths, to the cause of a higher percentage of dark moths in the population.
2. All moths have equal status when it comes to mating or being eaten by birds.
3. A moth does NOT have the intention to mate with one type of moth over the others.
4. A moth does NOT reproduce offspring with favorable traits on purpose. Therefore, the color traits of the moth offspring can be random.
5. The specific local interactions of the moths cannot directly affect the changing pattern”

Instruction on this implication feature was given after introducing the “converging change” inter-level feature.

To illustrate the other implication feature of “not align or non-matching between agents and the pattern” (Feature 7b, Table 1), students were asked to observe a data table and explain why it occurred that the number of dark-colored moths dropped at Generation 5 from 229 to 220 while the percentage of dark moths increased over time from 61% to 67%. This prompt question appeared after students learned the “collective summing” feature.

The possible answer stated that the pattern of moths getting darker reflected all the moths’ interactions (i.e., interacting with other moths, with birds, and with the environment). The answer further explained that it was possible for the number of dark moths to decrease in some generations, but white moths declined more than dark moths.

Therefore, the percentage of dark moths still increased over time. In other words, moths' local interactions did not always align with the overall pattern. After giving the reason, the instruction text explicitly mentions "not always align or non-matching is one of the inter-level implication features in the PAIR-C framework. To be specific, agent-level properties (i.e., moths of different body colors being born) do not always align with the overall pattern (i.e., the increasing percentage of darker-colored moths)."

In the first Peppered Moths module, all seven PAIR-C features were integrated into the instruction for the PAIR-C intervention group. Moreover, detailed instructions were provided to students which explain the relationships between the external agent and the internal agent in the mice scenario. For example, an instructional text stated "All mice despite fur colors can be spotted and eaten by birds. Birds preying on mice are more likely to eat the most visible white mice, but they can eat darker mice when they happen to be around and are seen. Similarly, some white mice who are not seen by birds can also survive. (PAIR-C Feature 1: Same/Uniform set of interactions)."

For the second pocket mice module, all seven PAIR-C features were illustrated, and instructions were given on how to apply the PAIR-C features in understanding and explaining the emergent process of natural selection in the context of rock pocket mice.

To be specific, after watching a link-visible ABM video showing dark-colored mice becoming more common, students were asked to describe their observations using a PAIR-C lens. The possible answer mentioned the four interaction features and provided an example to illustrate each feature ("All mice have the same set of interactions available to them. For example, both dark-colored mice and light-colored mice could get eaten by owls, and those mice who survive the predators can mate and reproduce offspring when they reach adulthood. Feature 1: Same/uniform set of interactions”).

When illustrating the inter-level features using the mice model, the researcher captured the converging change process in another link-visible simulation video showing all interactions (mice mating interactions and bird-mice chasing interactions) approximate the final pattern of dark-moths becoming more common over time. Following the video, instructional texts further stated that “One cannot predict the exact number of mice surviving and reproducing in each generation, but it is possible to predict that dark-colored mice are spotted and eaten by birds less frequently in the darker environment, thereby having more chances to survive and reproduce more offspring than light-colored mice. Over time, mice with a favorable dark-colored trait that survive in the dark environment will reproduce and become more popular in the population.”

In sum, the two simulation scenarios designed for the PAIR-C intervention group integrated all the PAIR-C features by modifying original ABMs, generating MERs from the modified ABMs, and providing instructional content to explain the PAIR-C features and demonstrate the use of these features in describing and reasoning about the emergent process. In the second mice scenario, a review of all the PAIR-C features and their applications in explaining the emergent process of natural selection was given to students near the end of the module.

The Regular Module

Unlike the PAIR-C module which stated that students would learn about the PAIR-C framework and how the PAIR-C features were relevant to understanding natural selection, the Regular module focused on the following learning objectives: 1) explain how the color of moths/mice changed over time; 2) investigate the concept of inheritance of a trait across generations; 3) learn about how natural selection affects a population of moths/mice over time; 4) use the simulation to design and perform your own experiments on natural selection.

In terms of the simulation models, both the Moth and the Mice Regular modules did not modify the original models by including "links". Hence, MERs used in the Regular module did not show "links" to visualize agent interactions. When describing observations as the regular ABM simulation ran, instructions focused on describing a set of rules that govern the behaviors of individual computational agents without explicitly showing the interactions. For example, after showing the possible answer that when there was no pollution in the environment, light-colored moths became more common, and dark-colored moths became less common over time. Instructional texts explained how the result was produced in the simulation: "The programmer who built this simulation came up with the following rules when developing the simulation:

1. each individual moth moves randomly in the simulated world
2. moths can mate with moths that are close to them
3. moths can give birth to other moths of different colors
4. moths can get eaten by birds or other predators in the space
5. moths' body color helps them blend in or stand out in their environment to the birds.
6. moths die a lot faster when the moth population exceeds the available resources."

In addition, more instructions were given to help students align their observations with two Darwinian principles:

- Random intra-species variability (individual variation): individuals in one generation of a particular species differ from each other in a number of dimensions, including physical characteristics (size, color) mental characteristics (perception, memory, intelligence), and behavioral patterns (child-rearing, feeding).
- Heritability of certain traits (genetic determination): some dimensions of variation are genetically determined, and other dimensions of variation are acquired (i.e. individual values reflect experience and lifestyle). Genetically determined characteristics are generally considered to be relevant for evolution.

Students in the Control group later received standard instruction on all of the five Darwinian principles based on their interactions with the models and the multiple representations. For instance, after students investigated how the distribution change of color variation was influenced by adjusting the selection pressure parameter, the instructional text asked students to notice that “The survival and reproduction rates of individual moths affect the moth species. Specifically, when there is no pollution, light-colored moths have higher survival and reproduction rates compared to other types of moths because the living environment is clean with light-colored trees. Under this condition, when the selection factor is high and the trees are a lit color, lighter-colored moths are better camouflaged than darker-colored moths so they survive and reproduce at a higher rate over time.” Furthermore, when prompted to explain why dark moths becoming more common, a complete answer which applied the two Darwinian principles of differential survival and reproductive rates was given to students “As pollution level increases, dark-colored moths happen to be camouflaged in the soot-filled environment and avoid being spotted by birds. Therefore, dark-colored mice have higher survival and reproduction rates compared to other types of mice."

Another unique instructional approach used in the Control group was the explicit mention of common misconceptions in the instructional texts. In response to the same prompt question that asked students to explain why Tick 0 looked different from Tick 125, the possible answer started with common misconceptions "When asked to explain why the changes happen over time (why tick 0 is different from tick 125), people would typically confuse color variation changes as an intentional behavior. For instance, people may say that there are more dark-colored moths because light-colored moths want to change to darker body colors to better camouflage and escape from birds in the polluted sooting environment." Followed by the common misconception, instructional texts emphasized that "There was **NO moth who could control or cause color variations.** In the meantime, although environmental conditions change (i.e., cycling pollution) over time, we **cannot attribute environmental factors as the direct cause** for changes in moth color variations over time.

The Regular module also adopted instructional text gathered from multiple sources (i.e., pilot study learning materials, available online resources, and literature on teaching and learning natural selection) reviewed by a content expert. For instance, when explaining how the variation in the mice population originated in the real world, the instructional text stated that "Genetic variations can arise from gene variants (also called mutations) or from a normal process in which genetic material is rearranged as a cell is getting ready to divide (known as genetic recombination). Genetic variations that alter gene activity or protein function can introduce different traits in an organism. Moreover, the initial genetic variations occur in the population through mutations during DNA Replication. Mistakes and random errors made during DNA replications result in different phenotypes (physical attributes) and changes in trait variations in the mice population across generations."

Furthermore, the Regular module provided more detailed instructions on explaining the functionality of certain simulation parameters. For example, when being introduced to the use of the selection slider, the instructional text for the control group explained "The SELECTION slider determines how moths are harvested by the birds that feed on them. In this simulation, SELECTION wraps up nicely many factors that determine the survivability of a species - how many birds there are, how hungry they are, and just how important camouflage is to escaping predation. SELECTION provides a probabilistic window - the lower the level of the slider, the wider this window." This information was not given to students in the intervention group.

In sum, for the two simulation scenarios, the design of the Regular modules did not modify any original ABMs. MERs were directly generated from the original ABMs without showing visual cues such as the links. All five Darwinian principles were illustrated, and instructions were given on how to apply the five Darwinian principles in understanding and explaining natural selection in the context of peppered moths and rock pocket mice.

Learning Activities

Each simulation scenario consisted of four activities, a total of eight activities for the two simulation scenarios. These activities were adopted from previous ABM integration practices for teaching natural selection and evolutionary processes (Wagh & Wilensky, 2017; Wilensky & Novak, 2010; Peel et al., 2019). The structure and the learning goals of the activities were similar across both groups. A detailed description of the first four activities used in the Peppered Moths module was elaborated below. In the description, I also elaborate on the similarities and differences between the PAIR-C module and the Regular module regarding what was shown and taught in each activity. After describing each activity, a box of prompts was listed with the intention to illustrate the instructional moves planned for facilitating each activity. Before the participants got into the specific activities, they were first introduced to the learning objectives and were provided with a background text on the specific simulation scenario. The background texts for the Peppered Moths scenario and the Rock Pocket Mice scenario can be found in Appendix L. After reading the text, students were asked to raise two questions that they would like to investigate in the following activities.

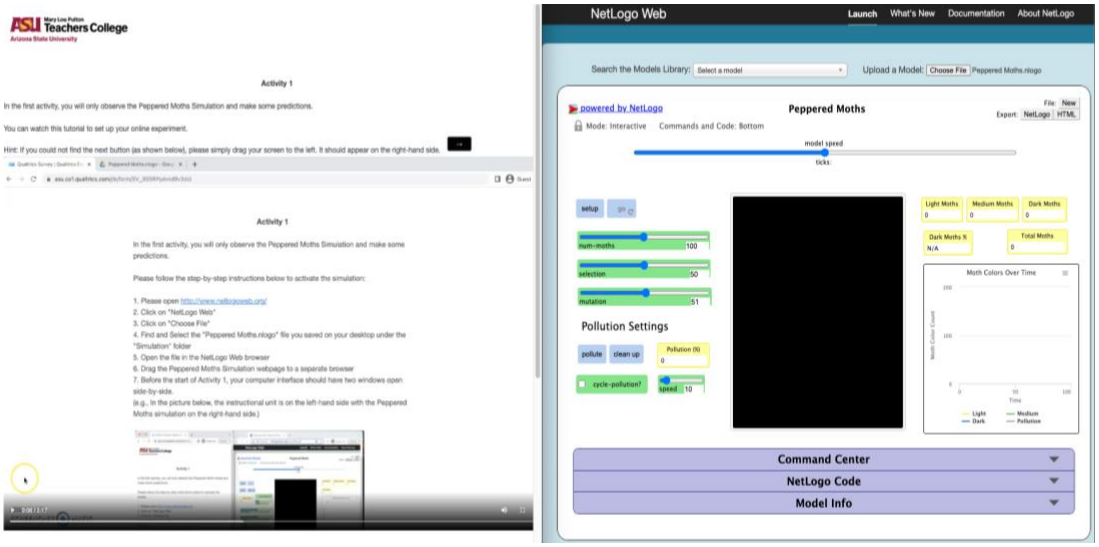
Activity 1

The first activity is intended to get students familiar with the simulation interface. Students in both conditions were provided with the same tutorial video and a step-by-step guide to set up the online module with the Peppered Moths scenario. Figure 8 showed how the screen should look at the initial state of the online learning environment.

After an initial description of the simulation environment, the same video tutorial was given to both conditions which focused on displaying all key components of the Peppered Moths simulation interface, including the “simulation world”, “buttons”, “switch”, “sliders”, “model speed controller”, “monitors”, “line graph”. The first activity closed by having students predict how the percentage of moths with each color would change over time without starting the simulation.

Figure 8.

Initial State of the Online Learning Environment



Activity 1

In the first activity, you will only observe the Peppered Moths Simulation and make some predictions.

You can watch this tutorial to set up your online experiment.

Hint: If you could not find the next button (as shown below), please simply drag your screen to the left. It should appear on the right-hand side.

Activity 1

In the first activity, you will only observe the Peppered Moths Simulation and make some predictions.

Please follow the step-by-step instructions below to activate the simulation:

1. Please open http://www.netlogoweb.org/
2. Click on "NetLogo Web"
3. Click on "Choose File"
4. Find and Select the "Peppered Moths.nlogo" file you saved on your desktop under the "Simulation" folder
5. Open the file in the NetLogo Web browser
6. Drag the Peppered Moths Simulation webpage to a separate browser
7. Before the start of Activity 1, your computer interface should have two windows open side-by-side.

(e.g., In the picture below, the instructional unit is on the left-hand side with the Peppered Moths simulation on the right-hand side.)

Search the Models Library: Select a model

Upload a Model: Choose File Peppered Moths.nlogo

Mode: Interactive

Commands and Code: Bottom

num-moths 100

selection 50

mutation 51

Pollution (%) 0

speed 10

Light Moths 0

Medium Moths 0

Dark Moths 0

Dark Moths % N/A

Total Moths 0

Note. In this picture, the instructional unit is on the left-hand side with the Peppered Moths simulation on the right-hand side.

Box 1

Activity 1 Prompt 1. *Press the "Setup" button on the simulation.* ***Describe*** *what you see in this simulation environment.* 2. *Before running the simulation, please* ***predict*** *how the percentage of moths with each color will change over time under the current circumstance.*

Activity 2

The second activity allowed students to run the simulation with the purpose of understanding how variations within a species change over time because of selection pressures. For the PAIR-C intervention group, students learned to use the PAIR-C four interaction features to understand and describe the emergent process of natural selection. For the Regular control group, students learned about how selection pressures connect to Darwinian Principles.

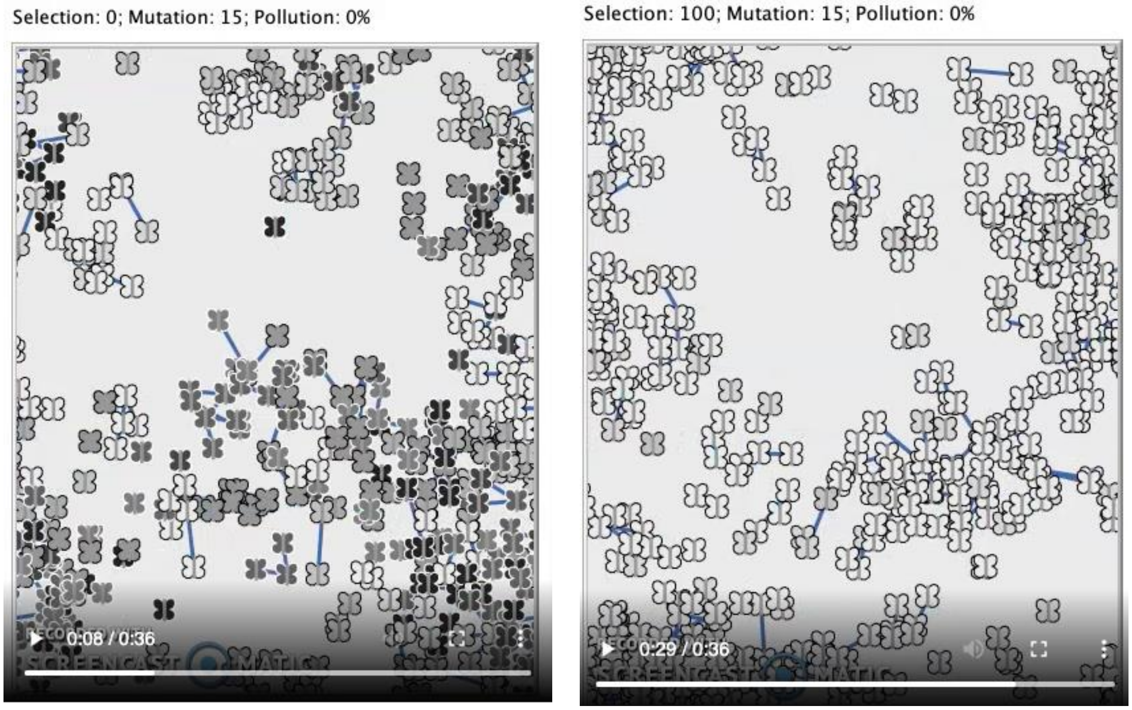
To start with, students first run the simulation according to the instructions and describe what they observed in the simulation world from tick 1 to tick 100. This gave students an opportunity to check their predictions from Activity 1. After the description, students in the intervention group received a possible answer with instructions on the PAIR-C framework. The instructional text defined the Pattern, Agents, and Interactions dimensions of PAIR-C and pointed out that the pattern-level behavior was governed by the rules of interactions at the agent level in this simulation scenario. Students in the control group received a possible answer with explanations about how the simulation was built by following rules developed by the programmer. Two Darwinian principles that aligned with their observations were introduced to students.

A multiple-choice question with three diagrams that showed a different distribution of moth color variations was given to students in both groups to check their understanding. This question could be regarded as the first knowledge-checking point in students' learning process. Upon completing this question, the intervention group received one version of the correct answer which emphasized that each moth agent followed the same rules to produce the changing pattern. The control group received another version of the correct answer which did not use any PAIR-C terms such as "changing pattern" or "agents". Again, students were prompted to reflect on their understanding after learning about the answer to this multiple-choice question.

Since students already had some initial interactions with the Moths model by observing the moth color changes over time, they were then introduced to the selection pressure parameter. Before manipulating the parameter, they first watched a simulation video recorded by the researcher showing how the moth color changed when other conditions remained the same (mutation 15, pollution 0%) while the selection pressure was at 0 and 100 respectively. Screenshots of the video were shown in Figure 9 below.

Figure 9.

Screenshots of the Simulation Video with Different Selection Pressure



Selection: 0; Mutation: 15; Pollution: 0%

Selection: 100; Mutation: 15; Pollution: 0%

Note. The screenshots were taken from the link-visible ABM video used in the PAIR-C intervention group. The video in the control group did not show any links.

Students in both groups were asked to describe their observations from watching the video (see prompt 5 in Box 2). Then, students in the intervention group were given an instruction page with information that illustrate the four interaction features (i.e., uniform, random, simultaneous, and independent). Students in the control group were given a different instruction page that focused on explaining the functionality of the selection slider presented in the simulation interface.

To close this activity, students in both groups were asked to experiment with the Moths model with different values placed on the selection slider, and then explain how the distribution change of color variation was influenced by selection pressures. After completing the task, possible answers and instructions were given to students. The intervention group received more instruction on using the PAIR-C interaction features to explain their observations whereas the control group received instruction on using Darwinian principles to explain the changes.

Box 2

Activity 2 Prompts

1. *Observe the simulation and describe what you observed in the simulation world from tick 1 to tick 100.*
2. *What information on this page did you miss in your response?*
3. ***Select the diagram below that best describes what will happen in terms of the distribution of color variations among moths over 100 ticks?***
4. *What information on this page is most important for you to understand?*
5. *Watch a simulation video. Describe what you notice when the selection pressure is at 0 versus the selection pressure is at 100.*
6. *Run the simulation and explain how the distribution change of color variations is influenced by selection pressures. Try setting at least three different values of the selection slider (e.g., the selection at 10, 50, and 90 respectively).*
7. *What information on this page is most important for you to understand?*

Activity 3

The third activity guided students to explore the pollution settings of the simulation with the purpose of examining and understanding the role of the environment in natural selection. It started off by having students set the pollution level to be around 50% while controlling for the selection parameter to be at 30. Students in both groups described their observations first and then watched a simulation video showing the same changing patterns of medium-colored moths becoming more commonly captured by a researcher. The intervention group watched the video with links presented. Instruction was accompanied by the explicit mention of the four interaction features that were perceivable from the link-visible ABM video. In contrast, the control group watched the video without links shown. Instruction focused on describing the random movement of the moths, and the distribution of different colored moths at the initial and the final state of the simulation.

Then, students in both groups were asked to continuously click the “pollute” button and set the pollution level to around 80%. Students run the simulation till tick 200 and describe their observations. Similarly, a simulation video captured by the researcher was provided to students showing the same changing pattern. A prompt question then asked students to share what they learned from the simulation video. This question could be regarded as a knowledge-checking question since students could answer it by applying what they learned from the instruction provided for the previous simulation video. Specifically, students in the PAIR-C intervention group were expected to include the four interaction features in their responses whereas the control group was expected to notice the randomness of moths movement as well as differences between the initial state and final state of the simulation.

As students were familiar with the single-direction pollution condition, they were introduced to the “cycle-pollution” switch which would cyclically change the environment from light pollution to high pollution while running the simulation. Students were also told to control for other settings, such as making sure the “selection” slider was at 30, the speed for pollution was at 10, and there were at least 60 moths in the simulation world. Next, students in both groups engaged in a data collection task that directed them to run and pause the newly configured simulation at six different time points (tick 0, tick 25, tick 50, tick 75, tick 100, tick 125). Students were also asked to take and upload a screenshot for the simulation at tick 0 and tick 125 respectively. Upon completion, students were asked to explain how the color variations at Tick 125 were different from Tick 0. The PAIR-C intervention group received a possible answer explaining color variations with six ABM screenshots representing the changing pattern over time. In contrast, the control group focused on the two-time points and thereby showed two screenshots at Tick 0 and Tick 125 with explanatory texts.

To understand how the pattern transformed, students were asked to explain why Tick 0 looked different from Tick 125. Instructions for the intervention group introduced the inter-level PAIR-C feature of “converging change” and provided an elaborative answer which considered interactions of all the agents when explaining the changing pattern. Instruction for the control group pointed out a common misconception for explaining why the changes happened over time and then also provided an elaborative explanation that applied Darwinian principles. The control group also mentioned that although environmental conditions changed (i.e., cycling pollution) over time, one could NOT attribute environmental factors as the direct cause for changes in moth color variations over time.

Box 3

Activity 3 Prompts

1. *Reset the simulation ("selection" slider to 30, pollution level to be around 50%). Run the simulation till around tick 100. Describe what is happening in the simulation world.*
2. *What information on this page did you miss in your response?*
3. *Do NOT reset the simulation. Continue clicking the "pollute" button to set the pollution level at 80%. Now, run the simulation till tick 200 and describe what you observed.*
4. ***What do you learn from this simulation video?***
5. *Explain how the color variations at Tick 125 is different from Tick 0.*
6. *Explain why Tick 0 looks different from Tick 125?*
7. *What information on this page is most important for you to understand?*

Activity 4

In this activity, students in both groups had more hands-on experiences by recording data from the model and examining how mutation might result in changes in variation distribution by manipulating the mutation slider. Despite the visual differences of showing or not showing links in the model world, the PAIR-C module gave instructions on the inter-level PAIR-C feature of "collective summing" and the implication feature of "not align". In contrast, the Regular module followed standard instructions by explaining how the mutation slider worked in the model and how Darwinian principles could be used to explain natural selection when considering the chance of mutations.

At the beginning of the activity, students in both groups followed step-by-step instructions to reset the simulation (i.e., Step 1: Turn off the "cycle-pollution" switch and set the pollution level to be at 90%. Step 2: Make sure the "selection" slider is set to 50, and there are at least 100 moths in the world. Step 3: Set the "mutation" slider to 50.)

Then, students were asked to pay attention to the percentage (%) of dark moths at tick 0 by observing the monitor boxes shown in the simulation interface. Students were instructed to pause the simulation at five different time intervals and record the % of the dark-colored moths for each time interval in a box.

This data collection practice allowed students in both groups to notice that the % of dark-colored moths did not always increase although the overall pattern showed that dark-colored moths were more common in the population. Students were later prompted to decide between True or False for a statement (see prompt 1 in box 4) which should be an easy answer given their experience from data collection. To explain why a newborn moth would not always be darker than its parents, the intervention group received instruction on the PAIR-C implication feature of "no direct effect" whereas the control group was instructed on the mutation mechanisms.

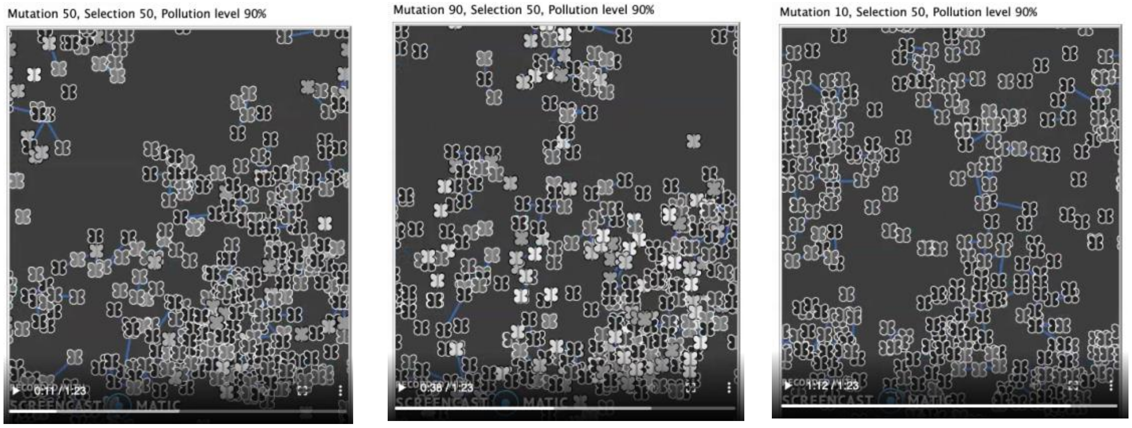
Next, students in both groups were provided with a simulation video that showed moths' color changes under different mutation levels while controlling for selection pressure, pollution level, and the number of moths (See Figure 10).

Students in both groups were asked to explain how the pattern changes over time from watching the video (see prompt 3 in Box 4). Then, students in the intervention group were given an instruction page with information that illustrates the inter-level feature of "collective summing". Students in the control group were given a different instruction page that focused on explaining the influence of the mutation slider and how Darwinian principles could be applied in interpreting the resulting changes.

The last task of this activity for both groups was to observe a data table and provide explanations for a prompt question (prompt 5, Box 4). Upon responding to this question, the intervention group was introduced to the implication feature of "not align" whereas the control groups were shown how to use Darwinian principles to understand the phenomena.

Figure 10.

Screenshots of the Simulation Video with Different Mutation Levels



Note. The screenshots were taken from the link-visible ABM video used in the PAIR-C intervention group. The video in the control group did not show any links.

Box 4

Activity 4 Prompts

1. *Decide True or False for the following statement: If a newborn moth has a mutation, its color will always be darker than its parents. Please explain your response in 2-3 sentences.*
2. *What information on this page is most important for you to understand?*
3. *Explain below how the pattern changes over time.*
4. *What information on this page is most important for you to understand?*
5. *Observe the data table provided by a researcher. Please explain why it occurs that the number of dark-colored moths drops at Generation 5 from 229 to 220 while the percentage of dark moths increases over time from 61% to 67%.*

Note that in Box 1 to Box 4, those underlined prompt questions were reflection questions that always appeared after an instructional page with possible answers. Students in both groups were prompted to reflect on what they learned from the instructional text throughout the activities as a way of showing cognitive engagement and mental processing with the instructional text.

All the simulation learning activities can be found in Appendix L.

CHAPTER 6

DEEPER UNDERSTANDING OF NATURAL SELECTION

Introduction

Because one of the research questions of the study is to investigate the effects of the PAIR-C module versus the Regular module on fostering students' deep understanding of Natural Selection and knowledge transfer, examining the learning outcomes related to the science content is essential to my investigation. In this chapter, I present the findings on five aspects of students' understanding of Natural Selection and transfer concepts. These aspects include: 1) students' overall understanding of natural selection measured by pre-posttest questions, 2) students' deep understandings of natural selection measured by pre-posttest deep questions, 3) students' ability to explain the process of natural selection in context transfer questions, 4) students' self-reported changes in understanding and explaining natural selection during the post-project interview, and 5) students' understandings of other emergent phenomena measured by pre-posttest far transfer questions.

Findings

Students' Overall Understandings of Natural Selection

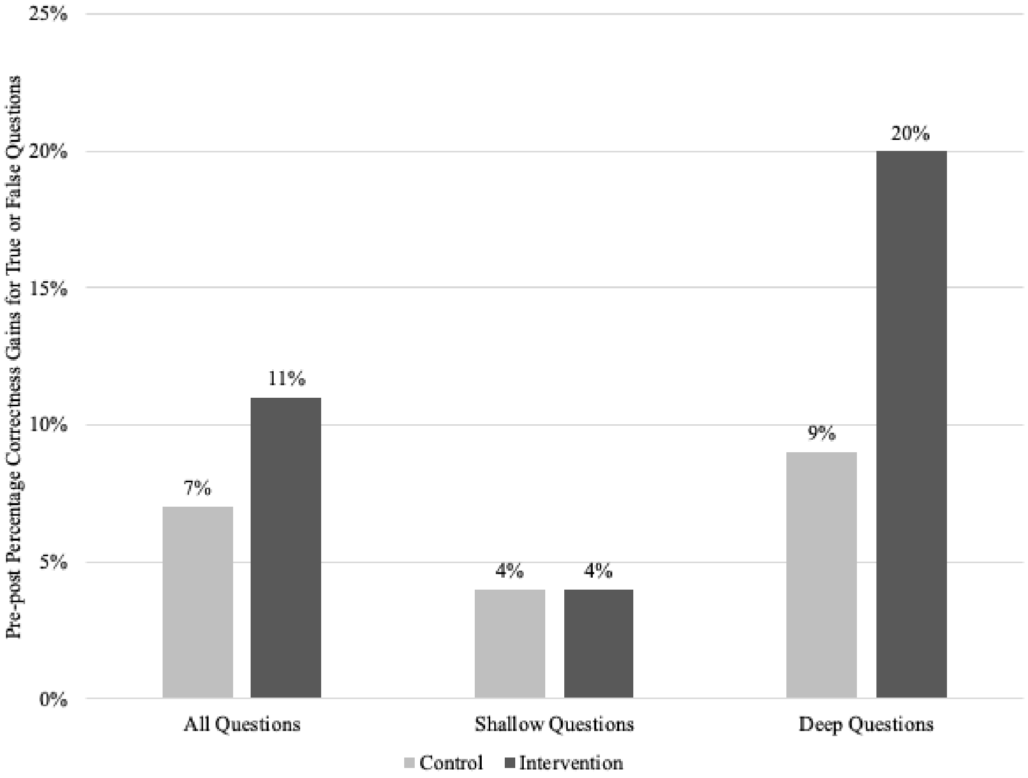
For measuring the overall understanding of natural selection, students' performance on the three sets of True/False natural selection questions was used in the analysis. Specifically, this analysis contained a total of 15 first-tier T/F items and the corresponding open-ended explanations for the false items. Because 10 out of the 15 items were false statements that needed explanations from students in their second-tier responses, students could receive a total of 10 points if they correctly identified and explained all the false items. Therefore, a maximum of 25 points could be earned by students in the knowledge test that assessed understanding of natural selection concept.

Results show that student performance on the T/F items improved significantly from the pretest to the posttest for both groups. The control group (N =26) improved from a mean score of 8.94 (SD = 3.09) to a mean score of 10.89 (SD = 2.87), t =2.60, p = .016, with a medium effect size d = 0.51. The intervention group (N = 24) improved from a mean score of 9.31 (SD = 2.78) to a mean score of 12.73 (SD = 4.36), t =4.42, p < .001, with a high effect size d = 0.90. These results indicate that both the Regular modules and the PAIR-C modules used in this experiment could successfully improve students' overall understanding of natural selection.

Given my interest was to investigate group differences attributed to the different module treatments, I conducted ANCOVA by using pretest scores as covariates. The result shows that there was a marginally significant difference between the two groups in the overall understanding of natural selection (F = 2.94, p = .093, partial η^2 = .059). This difference is also reflected in the first two columns of Figure 11. As shown in the figure, the percentage of correctness gains was 7% for the control group whereas the percentage of gains was 11% for the intervention group. This result indicates that the intervention group who used the PAIR-C modules had a better performance on the overall understanding of natural selection compared to the control group who used the Regular modules.

Figure 11.

Pre-post Percentage Correctness Gains for All, Shallow, and Deep T/F Questions



Students' Deep Understandings of Natural Selection

To measure students' deep understanding of natural selection, I focused on students' performance on the deep questions of the pre-posttest. I also discriminated students' performance between answering the deep questions and the shallow questions. Table 7 shows the descriptive statistics for students' performance on the pre-posttest deep and shallow questions by comparing the group mean score for each T/F item. Note that the mean score was based on the first-tier T/F question.

Table 7.

Comparison of Group Mean for Each Item

Item No.	Q1a (Deep)		Q1b		Q1c		Q1d		Q1e	
Condition	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.
Pre Mean	.31	.08	.62	.67	.85	.96	.58	.75	.96	.87
Post Mean	.04	.25	.62	.46	1.00	1.00	.88	.92	1.00	.92
Trend	⬇	⬆	=	⬇	⬆	⬆	⬆	⬆	⬆	⬆
Item No.	Q2a (Deep)		Q2b (Deep)		Q2c (Deep)		Q2d (Deep)		Q2e	
Condition	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.
Pre Mean	.12	.21	.19	.21	.08	.08	.15	.08	.96	1.00
Post Mean	.27	.29	.38	.58	.04	.13	.00	.13	.96	1.00
Trend	⬆	⬆	⬆	⬆	⬇	⬆	⬇	⬆	=	=
Item No.	Q3a (Deep)		Q3b		Q3c (Deep)		Q3d (Deep)		Q3e	
Condition	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.	Ctrl.	Intv.
Pre Mean	.27	.42	.92	.79	.50	.33	.00	.08	.96	.96
Post Mean	.69	.79	.65	.79	.62	.58	.31	.38	1.00	.87
Trend	⬆	⬆	⬇	=	⬆	⬆	⬆	⬆	⬆	⬇

From Table 7 (also in Appendix C), eight out of fifteen T/F items were labeled as deep questions that assessed understanding of the causal mechanisms for natural selection. Meanwhile, the rest of the seven items were labeled as shallow questions which assessed basic understandings of Darwinian principles, knowledge of either agent-level or pattern-level understanding, or implications. Note that deep questions were items with a mean score equal to or lower than .50 in the pretest whereas shallow questions were items with a mean score higher than .50 in the pretest.

Moreover, all eight deep questions were false statements whereas only two shallow questions were false which needed students to explain the reasons in the second-tier responses. Therefore, students could receive a maximum of 16 points for answering and explaining 8 two-tiered deep questions, and 9 points for answering 7 shallow questions and explaining 2 false statements. Table 8 below shows examples of a shallow question and a deep question to illustrate the differences.

Table 8.

Examples of Deep and Shallow questions

Question Level	True or False Item Statement	Pretest Correct%
Shallow Question	**Q1d.** The gray mice chose not to reproduce with the brown mice and not to give birth to brown mice because the gray fur color trait was more advantageous.	Ctrl: 58% Intv: 75%
Deep Question	**Q1a.** The number of gray mice in the population increases slightly each year for 20 years, which adds up to the pattern of gray fur becoming more common in the population.	Ctrl: 31% Intv: 8%

Q1d was a shallow question because it focused on the agent-level interactions (e.g., whether the gray mice chose to mate with brown mice) and did not require an understanding of how the mice agents' interactions produce the pattern of fur color change over generations. There were 58% of the control group participants and 75% of the intervention group participants answered this question correctly in the pretest. **Q1a** was considered a deep question because it required a deeper understanding of the inter-level causal relationships. A correct causal understanding of this statement would imply that the increase of gray mice was not gradual or continuous by every generation. One year there could be a surge of gray mice and one year there could be many dying instead. The pattern of gray mice becoming more common in the population was the net effect of adding positive and negative numbers of gray mice considering all the interactions within that generation and then comparing across generations. There were 30% of the control group participants answered this question correctly in the initial pretest. Only 8% of the intervention group participants correctly determined this statement as false.

Results show that students' deep understanding improved significantly from the pretest to the posttest for both groups. The control group improved from a mean score of 2.54 (*SD* = 2.16) to a mean score of 3.73 (SD = 2.52), t =2.08, p = .048, with a medium effect size d = 0.41. The intervention group improved from a mean score of 2.52 (*SD* = 2.15) to a mean score of 5.56 (*SD* = 3.95), t =3.90, p < .001, with a high effect size d = 0.80. These results indicate that both the Regular modules and the PAIR-C modules used in this experiment could successfully improve students' deep understanding of natural selection.

Again, since I was more interested in investigating group differences attributed to the different module treatments, I conducted ANCOVA by using pretest deep question scores as covariates. The result shows that there was a statistically significant difference between the two groups in answering the deep questions of natural selection (F = 4.16, p = .047, partial η^2 = .081). This difference is reflected in the last two columns of Figure

11. As shown in the figure, the percentage of correctness gains for answering deep questions was 9% for the control group whereas the percentage of correctness gains was 20% for the intervention group. These results indicate that the intervention group had a significantly better performance in answering deep questions of natural selection compared to the control group after the modules. In other words, the PAIR-C modules were more effective in fostering students' deeper understanding of natural selection compared to the Regular modules. In contrast, students' performance did not differ between groups in answering shallow questions ($F = .01$, $p = .934$, partial $\eta^2 = .000$). This pattern is shown in the middle two columns of Figure 11.

Students' Abilities to Explain the Process of Natural Selection

To see whether the PAIR-C module could further impact students' abilities to explain the process of natural selection, I conducted analyses based on their responses to the two context transfer questions and one reflective prompt question. The two context transfer questions (See Q4 and Q5 in Appendix A) were essentially the same in terms of measuring how students explain the process of natural selection in the context of living mouse species evolving with claws and living mosquito species evolving resistance to DDT. The reflective prompt question (See prompt 39 in Appendix D) appeared near the end of the second module which asked students to reflect on the important information presented on the instructional page. The instructional page provided a five-step explanation for the natural selection process in the context of living mouse species evolved with favorable traits. It is assumed that how students understand and reflect on the five-step explanations of natural selection during the module would influence their performances in answering the two contextual questions in the post-test. Therefore, the analyses of these questions could inform students' abilities to explain the process of natural selection.

Results show that there was no significant improvement in answering the two context transfer questions from the pretest to the posttest for both groups. The control group improved from a mean score of 1.73 (SD = 1.34) to a mean score of 1.77 (SD = 1.75), $t = 0.09$, $p = .928$, with a small effect size $d = 0.02$. The intervention group improved from a mean score of 2.04 (SD = 1.68) to a mean score of 2.38 (SD = 2.36), $t = 0.78$, $p = .445$, with a small effect size d = 0.16. For testing group differences in answering the post-test context transfer questions, ANCOVA was conducted. There was no statistically significant difference between the two groups ($F = 0.70$, $p = .409$, partial $\eta^2 = .015$). Similar results were also manifested in students' responses to the simulation prompt question. An independent sampled t-test showed that there was no significant difference between the two groups in responding to the reflective prompt question (M_{ctrl}= 0.81, M_{intv}= 0.87, $t = 0.33$, p = .746, d = 0.09).

As expected, a strong positive correlation was found between students' performance on the prompt question and their pre-post learning gains on the context transfer questions ($r = .615$, p < .001). This finding confirmed the previous assumption and indicated that when students could not understand or reflect on the five-step explanations of natural selection during the module, it would highly affect their performances in answering the two context transfer questions in the posttest. To further investigate why instruction on the natural selection process was not effective in helping students improve their abilities to explain the process across different contexts, the following analyses focused on what happened in students' actual learning processes.

When examining the responses to the simulation prompt question, students in both groups showed similar response patterns. Because prompt 38 asked students to re-organize the order of five statements in explaining the natural selection process, I scored their responses by counting the number in the correct order. A total of five scores could be earned if students put all the statements in the correct order. As a result, no student earned a score higher than 4 across groups. There were 46% of students from the Control group and 44% of students from the Intervention group failed to earn any points, which showed a lack of knowledge in explaining the process of natural selection for both groups. After re-organizing these statements, students received a possible answer on the following instructional page. For the Control group who used the Regular module, the answer included the correct order of the five statements and more instructions on using three Darwinian principles (i.e., DP3, DP4, & DP5) to elaborate the natural selection process. For the Intervention group that used the PAIR-C module, the answer also included the correct order of the five statements and more instructions. However, the instructions referred to only one PAIR-C interaction feature (i.e., Feature 1: same/uniform) to elaborate the process.

In their responses to the reflective prompt question (prompt 39) which asked them to reflect on the important information presented on the instructional page, students in both groups demonstrated similar patterns in terms of their reflections. One pattern was that a large proportion of students in both groups realized that the environmental condition or selective pressure could NOT influence an individual's trait because it was genetically determined (42% in the control group; 52% in the intervention group). Many students recognized that trait variation in the mice population was present prior to the specific selection pressure. Although this reflection was important in describing the process of natural selection, it did not further extend to the ability to explain the natural selection process across contexts.

Another pattern shown in students' responses was that most of the students in both groups did not attend to instructions beyond explaining the correct order. In the control group, only two students reflected on the importance of using Darwinian principles to elaborate on the process of natural selection. Unfortunately, none of the students in the intervention group emphasized the importance of the PAIR-C interaction features in understanding the natural selection process.

Students' Self-reported Changes in Explaining Natural Selection

Four students volunteered to participate in a 20-minute interview after completing the entire project. Two of them (hereafter Student A and B) were from the Control group and two other students (hereafter Student C and D) were from the Intervention group. During the semi-structured interview, students were asked to reflect on how their understanding of Natural Selection changed since the start of the project and how they would explain natural selection differently to others (See Q4, Appendix F).

Three general characteristics were identified from students' responses, including 1) self-identified common misconceptions of natural selection 2) tended to attribute their learning to the help of simulations in visualizing the process of natural selection, and 3) did not make an explicit connection to Darwinian Principles or PAIR-C features. Excerpts between the interviewer and the students are presented to illustrate these characteristics (Table 9).

Table 9.

Excerpts of Students' Responses

Excerpt	Students' Responses
Interviewer: In what ways, if any, has your understanding of Natural Selection changed since the start of the project?	
Excerpt 1 (From Control group)	Student A: Yeah, I haven't done, like a science class in quite some time. And when I did, you know, so I had a general overview of natural selection. But I realized that I *had some of the misconceptions* that we talked about. For example, when a species acquires a mutation, in my mind it was like because they're like choosing like I need to have long claws. But it's mainly like environmental-based. And that was something that I didn't grasp until the simulation that really helped me put that into place.
Excerpt 2 (From Control group)	Student B: So the simulation project basically built my knowledge of this stuff from the beginning. So now I understand that *natural selection is not an active choice of these animals*, it is a natural consequence of changes in their environment. They do not, for example, choose to find mates with more advantageous features. They just kind of happen to have the population that survives and finds their mates that way.
Excerpt 3 (From PAIR-C group)	Student C: From the start of the project, I kind of knew natural selection, but not really. I just like took from what I kind of remembered from my biology class and I was like, Well, it's kind of like survival of the fittest or whatever. While going into the simulation and like learning more about it, I was like, Oh, there's a lot that goes into this. *It's not just like, they die off because they're weaker or whatever.* It's very much of like, it's like the role of where they are, like where they live and all of this stuff. And I was like, Well, that's kind of crazy.
Excerpt 4 (From PAIR-C group)	Student D: Well, I'll give you, if I may, the better example is the mice with the fur and the simulations helped me see visually the effects of the ground and how the ground affected the color of the fur and connection with the prey or the predator. That was great, right? It's one thing to read it. It's another to visually see and play it.
Interviewer: How would you explain natural selection differently to others?	
Excerpt 5	Student A: Yeah, for sure. Like how I said about how the environmental

(From Control group)		like the environment really affects natural selection and how over time, through many generations, mutations are acquired. So I would put more, I personally would put more emphasis on like the environment and how they don’t like choose specifically who they’re going to mate with based on like who or what type of the species survives and also like the predators who are going to prey on them.
Excerpt 6 (From PAIR-C group)	Student C:	I would explain it in the sense that It’s not just how they survive that plays a role, but it’s like the way they mate and all these other things: where they live and food resources and all these other stuffs. It’s not just, oh, they’re just trying to survive. It’s a whole concept.
Excerpt 7 (From PAIR-C group)	Student D:	So one thing I took away from the excerpts was with natural selection, there wasn’t only randomness. And I definitely appreciated how there was multiple factors, especially the breeding concept. It wasn’t like the subjects saw a matter of attraction. There was also a matter of locality. There was also a matter of habitability. You know, a brown mouse to another brown mouse would mate with another brown mouse mostly because they didn’t know there was another color mouse nearby. You know, that has to be a factor. So I think that visually helped me see that versus just assuming. You know, we can’t assume. Oh, how come a brown mice just didn’t look around for a darker colored mice, you know *you can’t assume that a subject will act like a human.*

In the above excerpts, all four students replied by spontaneously identifying several common misconceptions of natural selection which they used to possess but now could avoid them. Specifically, students stated “I realized that I had some of the misconceptions that we talked about. For example, when a species acquires a mutation, in my mind it was like because they’re *choosing* like *I need to* have long claws. But it’s mainly **environmental-based**. (Student A)”, “So now I understand that natural selection is *not an active choice* of these animals, it is a **natural consequence of changes in their environment**. They do not, for example, *choose to find mates* with more advantageous features. They just kind of happen to have the population that survives and finds their mates that way. (Student B)", "It's not just like, they *die off because they're weaker* or whatever. It's very much of like, **it's like the role of where they are, like where they live and all of this stuff**. (Student C)", "It wasn't like the subjects saw a matter of attraction. There was also a matter of locality. There was also a matter of habitability. ….. we can't assume. Oh, how come brown mice just didn't look around for darker-colored mice, you know you can't assume that *a subject will act like a human*. (Student D)". These statements showed that students could not only identify common misconceptions such as teleological conceptions, anthropomorphism, and absolute thinking but also point out that correct understandings should consider the role of the environment and other factors.

In addition, all four students attributed their learning to the use of simulations and appreciated how the module helped them in understanding natural selection as a process. Specifically, by saying "simulations helped me see visually the effects of the ground and how the ground affected the color of the fur and connection with the prey or the predator.", student D explained how the visual representations shown in the simulation helped him in understanding natural selection as a process through which organisms adapt to their environment and interact with each other and predators in the environment.

Finally, the two students from the control group did not explicitly mention the application of Darwinian principles when reflecting on their changes in understanding. Similarly, the two students from the intervention group also did not make explicit connections to any of the PAIR-C features.

Although students showed similar characteristics when replying to the interview questions, Student C and Student D from the intervention group appeared to adopt a complex system thinking perspective when asked to explain natural selection. For example, by saying “It’s not just how they survive that plays a role, but it’s like the way they mate and all these other things: where they live and food resources and all these other stuff. It’s not just, oh, they’re just trying to survive. It’s a whole concept. (Student C)” and “I definitely appreciated how there were multiple factors, especially the breeding concept. (Student D)”, both students emphasized the multiple factors involved in the process. The multi-factor thinking often led to non-linear cause-and-effect reasoning in system thinking, which would assume that the whole was greater than the sum of its parts. In contrast, although Student A and Student B noted the role of the environment, they seemed to be overly focused on the environmental parameters without mentioning the other factors and the interactions.

Students’ Understandings of Other Emergent Phenomena

To examine how well students were able to transfer their understanding of the emergent process of natural selection to explain other emergent phenomena, I conducted analyses on their responses to the pre-posttest far transfer questions (Q6 & Q7) as well as the post-project interview.

The pretest far transfer questions only contained 2 items that asked students to explain why geese maintained a V-pattern during flight and how one could stop fake news from spreading through the internet. In addition to the same two items, the post-test far transfer questions also provided students with two ABM simulation videos corresponding to the two scenarios and asked students two sub-questions to explain how the simulation can help them understand the process of geese flying into V-shape as well as the process of fake news spreading. Students' written responses to each question were scored based on a specific coding rubric. For Q6a and Q7a, students could receive a maximum score of nine. For Q6b and Q7b, students could receive a maximum score of six. Therefore, a maximum score of 15 could be obtained for all the posttest far transfer questions (see Appendix G for the scoring rubrics and students' sample responses).

Results show that students in both groups did not have significant improvement from the pretest to the post-test in answering the two identical far-transfer items (total score of six). The control group had the same mean score of 1.65 from the pretest to the post-test. The intervention group slightly improved from a mean score of 1.54 to a mean score of 1.58. An independent t-test was conducted to compare the group differences in responding to all four posttests far transfer items which showed no statistically significant difference ($M_{Intv}= 3.13$, $SD_{Intv}= 1.73$; $M_{ctrl}= 2.73$, $SD_{ctrl}= 1.43$; $t = .875$, $p = .386$). A similar result of no conditional difference was found in an ANCOVA analysis using pretest far transfer scores as covariates ($F = .98$, $p = .327$, partial $\eta^2 = .02$).

In the post-project interview, students were asked to reflect on whether learning the natural selection simulation scenarios could be applied to understanding other complex concepts and whether they noticed any commonalities between other emergent phenomena and the emergent process of natural selection (See Q6, Appendix F). Neither of the four interviewees showed a strong ability to transfer and apply what they learned in the natural selection module to explain other emergent phenomena. No explicit mention or reference to the PAIR-C features from students who experienced the PAIR-C intervention modules.

Although there was no significant difference between the two groups, interesting patterns regarding misconception expressions were found when scoring students' written responses. In the pretest of the two far transfer items, the coders noticed four misconceived responses from students in each group. For instance, a student from the intervention group stated, “I think the V-pattern stays maintained because the leader is in front and controls the destination”. Similarly, a student from the control group expressed “The V- pattern is still maintained because …… It is also a way they can just follow the leader of the flock”. By assuming one goose with a leadership role that directs or causes the V-pattern, students from both groups hold misconceptions about causal agents (Feature 7a, Table 1). A correct explanation of the V- pattern formation should recognize that the V-pattern is maintained not due to stronger geese who control the direction of the entire flock but simply by individual goose seeking a place to fly that requires the least effort, which is somewhat diagonally behind another bird. In the posttest of the four far transfer items, the coders noticed nine misconceived responses on controlling agents from students in the control group but only one such misconceived expression from students in the intervention group. This finding indicates that students in the PAIR-C intervention group might demonstrate fewer misconceptions about understanding far-transfer concepts than those in the Regular control group.

Discussion

The finding that the overall post-test score of students in both groups significantly outperformed their pretest scores suggested that both the PAIR-C modules and the Regular modules could improve students' overall understanding of natural selection.

More importantly, when comparing group differences in responding to deep questions that primed students to reason about the inter-level casual relationships or the causal mechanism of natural selection, students in the intervention group who used the PAIR-C modules showed significantly better performance than those in the control group who used the Regular modules. This indicates that the learning module integrated with the PAIR-C features could contribute to deeper learning of emergent process concepts, such as natural selection. Moreover, students' responses to the interview question that asked them to reflect on their changes in the understanding of natural selection also indicated that the intervention group students seemed to develop a system thinking perspective by considering the interactions of multiple factors in explaining the natural selection process. In contrast, students in the control group seemed to merely focus on the environmental factors in explaining the process.

Nevertheless, students in both groups were not able to demonstrate significant improvement in explaining the process of natural selection across different contexts because they did not receive enough instructions on how to use PAIR-C features or Darwinian Principles to elaborate on the process of natural selection. Given that students' learning and reflections during the module had a strong positive correlation with their performance on explaining the natural selection process in posttest context transfer questions, more explicit instructions should be provided to scaffold students learning. To be specific, when presenting the five-step process, more instructions should be given to students that explain each step statement using relevant PAIR-C features. See Table 10 for the revised instruction.

Table 10.

Revised Instruction on Explaining Natural Selection Process

Steps	PAIR-C Explanation
Step 1. Mice that survive to adulthood are likely to reproduce and pass on their traits to their offspring.	Mice can reproduce randomly with any other mice (dark, medium dark, or light ones) even though they passed a less beneficial trait to their offspring. (Feature 2: Random others)
Step 2. From reproduction, random mutations might occur, resulting in variation among offspring.	Dark-colored mice cannot intentionally choose to mate with another dark mouse and give birth only to dark-colored offspring. (Feature 7a: no causal agent)
Step 3. Mice with favorable or unfavorable traits who survived would reproduce and pass on those traits to their offspring.	Dark-colored mice and light-colored mice have the same set of interactions. They both can survive, reproduce, give birth to offspring, and get chased by predators. (Feature 1: same/uniform set)
Step 4. When selective pressure is introduced, mice with favorable traits for the environment would have more chances for survival compared to those with unfavorable traits.	Although dark-colored mice have more chances for survival, it doesn't mean that light-colored mice will stop mating with each other and wait for dark-colored mice. Mating interactions co-occur at the same time and are independent. (Feature 3&4: simultaneous, independent)
Step 5. Over time, the population is mainly composed of mice with favorable traits.	The pattern of mice getting darker over time reflects all the mice's interactions with each other and with the predators. Therefore, it is possible for the number of dark mice to decrease in some generations even though the proportion of dark mice increases over time. (Feature 5 & 6: Converging change; Collective summing)

Finally, students in both groups did not achieve a better understanding of their responses to the far transfer questions. Although it is evident that the intervention group students made some progress in understanding other emergent phenomena by stating fewer misconceptions, the progress was limited. In the post-project interview, participants in the intervention group were still unable to spontaneously apply PAIR-C features learned from the modules into explaining other emergent phenomena. Ideally, students from the PAIR-C intervention group should be able to develop a consistent mental model and use a set of emergent PAIR-C features to explain emergent process concepts after the intervention modules. However, there was no instruction provided to connect natural selection concepts with other emergent phenomena during the modules. As suggested by the pilot studies, future efforts could consider adding a concept-general process module to introduce PAIR-C features in other emergent phenomena before providing the concept-specific natural selection module. Or as demonstrated in pilot study 2, the future design of natural selection modules could also refer to other emergent phenomena and explicitly demonstrate the connections and commonalities between these emergent processes.

In sum, this chapter shows the effectiveness of the PAIR-C module in learning natural selection, especially in developing a deeper understanding of natural selection compared to Regular modules. Because students showed limited progress and differences in answering the contextual transfer and far transfer questions, it is important to further probe their mental models when constructing causal explanations. This led to exploring more deeply about students' misconceptions underlying those causal explanations. In the next chapter, I will focus on capturing the characteristics or identifying patterns of students' misconceptions in both the intervention group and the control group.

CHAPTER 7

CHARACTERISTICS OF STUDENT MISCONCEPTIONS

Introduction

Researchers (Chi et al., 2012; Gregory, 2009) have recognized that students demonstrated common misconceptions such as teleological perspectives and centralized deterministic mindsets in their reasoning and explanations about natural selection. Hence, a deep understanding of natural selection should adopt an unintentional and decentralized mindset. More importantly, students should be able to apply features from a collective causal structure to explain the inter-level causal relationships of natural selection.

In Chapter 6, findings show that students in both groups could spontaneously identify common misconceptions when explaining natural selection in the post-project interview. Moreover, in answering the posttest far transfer questions, students in the intervention group showed less frequency of misconceived expressions on controlling agents compared to those in the control group.

To further examine students' misconceptions of natural selection and investigate the similarities and differences between the two groups, this chapter focused on analyzing the categories and number of misconceptions in students' open-ended responses to reason and explain natural selection. This chapter presents findings on four aspects, which include: 1) students' misconceptions expressed in their open-ended responses to the pre-posttest natural selection questions, 2) students' abilities to correct misconceptions displayed in the module, 3) students' overall performance on misconceptions and its relationship with learning outcomes, and 4) students' characteristics of misconceptions analyzed using epistemic network analysis. The overarching goal of Chapter 7 is to address RQ3 by testing the potential of the PAIR-C module versus the Regular module on abating robust misconceptions of natural selection.

Findings

Students' Misconceptions in the Pre-posttest

Students' misconception expressions on natural selection were captured from their open-ended responses to the pre-posttest questions. In the pre-posttest, three sets of True/False natural selection questions asked students to give corresponding open-ended explanations for the items they identified as false. Moreover, the two context transfer questions could also reveal students' misconceptions in explaining the process of natural selection. Hence, the analyses were based on students' responses to all the open-ended natural selection questions.

As mentioned in Chapter 4, there were nine categories of misconceptions coded to encompass students' misconceptions as well as eight types of PAIR-C features coded to capture the underlying individualistic causal structure of students' statements. Because a student could have repetitive misconceptions under the same misconception category in responses to different questions or even the same question, the number of different misconception categories was used as a dependent variable (DV1) in addition to the total amount of misconceptions expressed by students (DV2). When analyzing from the PAIR-C perspective, students' statements were first coded into each PAIR-C feature and then separated into two-dimensional features: interlevel causal relationship (ICR) features (DV3) and PAIR-C interaction features (DV4). For both DV3 and DV4, I used the same approach as in DV1 to avoid repetitive counting of PAIR-C features for each student. Table 11 illustrates an example of how I coded one student's responses into different categories and how I transformed the open-ended data into the four DVs as measures of misconceptions.

Table 11.

A Coding Example for Misconceptions and PAIR-C Features

Three Statements from Student H	Question	Misconception	PAIR-C
Statement 1: Birds' beaks continue to grow (Feature 5)/ therefore even a smaller beak bird could evolve into a larger beak bird (MC7)/.	Pretest Q3	MC7: Transformat-ional views	Feature 5: incremental change
Statement 2: All species grow, develop, mutate and evolve as time goes on. /Throughout each generation newly adapting features and characteristics are formed to better fit into the environment in which they live. (MC3, Feature 7b)/Possibly mating with other types of species could also result in different features than the previous generation./	Pretest Q4	MC3: Directed variation	Feature 7b: alignment
Statement 3: Evolving over time and building up a tolerance from being subjected to the insecticide./ Much like bacteria when it meets with antibiotics can evolve and grow more resistant. (MC1, Feature 5)/ Humans can build up immune systems to better fight infections and allergies and so can other living organisms. (MC5, Feature 7a)/ The mosquitos have grown more tolerant to DDT as they have more contact with it (MC3, Feature 7b)/ and their internal systems fight it off or try to. (MC5) /	Pretest Q5	MC1: Intra-generational change MC3: Directed variation MC5: Anthropomorphism	Feature 5: incremental change Feature 7a: causal agent Feature 7b: alignment

In the example of student H, he generated three statements in response to three different questions in the pretest. For each statement, different misconception categories and PAIR-C features were assigned to specific idea units separated by "/". It is evident that this student showed repetitive misconceptions (MC5) within Statement 3 and also demonstrated the same categories of misconceptions (MC3) across Statement 2 and Statement 3. Similarly, the student showed the same underlying PAIR-C features (Feature 5 and Feature 7b) across three statements. Therefore, in transforming this coding result into DVs, the student showed four different categories of misconceptions (DV1 = 4); a total of five misconceptions (DV2 = 5); three PAIR-C ICR features (DV3 = 3); zero PAIR-C interaction features (DV4 = 0). The transformed data were then analyzed using descriptive and inferential statistics.

Table 12 presents the means and standard deviation of each misconception DV for both the control group and the intervention group. According to the table below, on average, students from both groups held more categories and expressed more misconceptions in the pretest compared to the corresponding performance on the posttest.

Table 12.

Average Students' Misconceptions in the Pre-Posttest.

	Control (N = 26)		Intervention (N = 24)	
Measures	Mean	SD	Mean	SD
Pretest Misconception Categories	1.92	1.41	2.13	1.73
Pretest Number of Misconceptions	2.27	1.71	2.33	2.01

Posttest Misconception Categories	1.46	1.14	1.04	1.00
Posttest Number of Misconceptions	1.73	1.56	1.13	1.12

To see if there was a statistically significant reduction of students’ misconceptions from the pretest to the posttest, paired sample t-tests were conducted. Results show that students in the control group did not show a significant reduction of misconception categories nor the total amount of misconception expressed from the pretest to the post-test ($t_{category\_ctrl} = -1.15$, $p = .261$, $d = -.23$; $t_{number\_ctrl} = -1.11$, $p = .277$, $d = -.22$). In contrast, students in the intervention group showed a significant reduction on both misconception categories and total numbers of misconceptions with medium effect sizes ($t_{category\_intv} = -2.78$, $p = .011$, $d = -.57$; $t_{number\_intv} = -2.64$, $p = .015$, $d = -.54$). These results indicate that students in the intervention group demonstrated drastically fewer categories and amount of misconceptions towards natural selection after learning the PAIR-C modules while students in the control group did not show such significant reductions.

To further probe the differences between the two groups, ANCOVAs were conducted using pretest misconception categories and pretest number of misconceptions as covariates respectively. Results show that there was no statistically significant difference between the two groups on misconception categories and numbers ($F_{category} = 1.77$, $p = .190$, partial $\eta^2 = .036$; $F_{number} = 2.39$, $p = .129$, partial $\eta^2 = .048$). This lack of conditional difference indicates that although a significant reduction of misconception categories and numbers was found within the PAIR-C intervention group, we could not say that the PAIR-C modules had on average a better effect on reducing misconception categories and numbers compared to the Regular modules. More fine-grained analysis should be applied to identify the differences between group individuals.

Given that students' statements were also coded based on the PAIR-C features, Table 13 presents the means and standard deviation for both groups on the two dimensions of the PAIR-C features: the PAIR-C ICR features and the PAIR-C interaction features. According to the table, on average, students from both groups showed more PAIR-C features in their responses to the pretest questions compared to the posttest. In addition, on average, students from the two groups demonstrated more ICR features than Interaction features in both the pretest and the posttest.

Table 13.

Average PAIR-C Features Demonstrated in Students' Responses on Pre-Posttest.

	Control (N = 26)		Intervention (N = 24)	
Measures	Mean	SD	Mean	SD
Pretest PAIR-C ICR Features	1.23	0.99	1.42	1.25
Pretest PAIR-C Interaction Features	0.46	0.65	0.54	0.66
Posttest PAIR-C ICR Features	1.12	1.03	0.83	0.87
Posttest PAIR-C Interaction Features	0.31	0.55	0.13	0.34

To see if students demonstrated significantly fewer PAIR-C features from the pretest to the posttest, paired sample t-tests were conducted. Results show that students in the control group did not show a significant reduction in both the ICR features and the interaction features from the pretest to the post-test ($t_{ICR\_ctrl} = -.38$, $p = .704$, $d = -.08$; $t_{interaction\_ctrl} = -1.07$, $p = .294$, $d = -.21$). In contrast, students in the intervention group showed a marginally significant reduction on the ICR features and significant reduction on the Interaction features ($t_{ICR\_intv} = - 1.94$, $p = .065$, $d = -.40$; $t_{interaction\_intv} = -2.63$, $p = .015$, $d = -.54$). These results indicate that students in the intervention group demonstrated fewer PAIR-C features from the individualistic causal structure in explaining the emergent process of natural selection after learning the PAIR-C modules while students in the control group did not show such decreases.

Again, to examine the differences between the average performance for the two groups, ANCOVAs were conducted using pretest PAIR-C ICR features and Interaction features as covariates respectively. Results show that there was no statistically significant difference between the two groups ($F_{ICR} = 1.01$, $p = .320$, partial $\eta^2 = .021$; $F_{Interaction} = 2.09$, $p = .155$, partial $\eta^2 = .043$). This lack of conditional difference echoes previous findings and indicates that there was not enough evidence to show the effectiveness of PAIR-C modules on reducing PAIR-C features of individualistic causal structure compared to the Regular modules by comparing group average results. More fine-grained analysis that could capture the characteristics of individual participants' misconceptions was needed to further identify the differences between groups.

Students' Abilities to Correct Misconceptions

Near the end of the second module, students were given an opportunity to select three misconceptions from a list of six and use what they learned from the modules to explain why the selected statements were misconceptions. The intervention group was primed to use the PAIR-C features to explain whereas the control group was prompted to use Darwinian principles. Table 14 shows the six common misconceptions. More details about this activity can be found in Appendix D (prompt 40) or Appendix L.

Table 14.

The Six Common Misconceptions

Misconception 1: Organisms change because they need to survive (Teleological, needs-based)
Misconception 2: Assign organisms with human feelings or actions (Anthropomorphism)
Misconception 3: Organisms develop a trait because they practice using it or they lose a trait because they do not use it (Use and disuse)
Misconception 4: Mutations arise because of selection pressure (Source versus sorting of variation)
Misconception 5: An entire population expresses the same traits, no variation in a population (Essentialist thinking)
Misconception 6: Natural Selection happens quickly and transforms the entire population with the goal of survival (Viewing natural selection as a discrete event)

When students successfully correct the misconception they selected, a score of "1" was given. Extra points were given to students who applied PAIR-C features or Darwinian principles in their explanations. A point of "1" was given to each correct application. Table 15 provides sample responses from students in both groups.

Table 15.

Sample Responses from Students

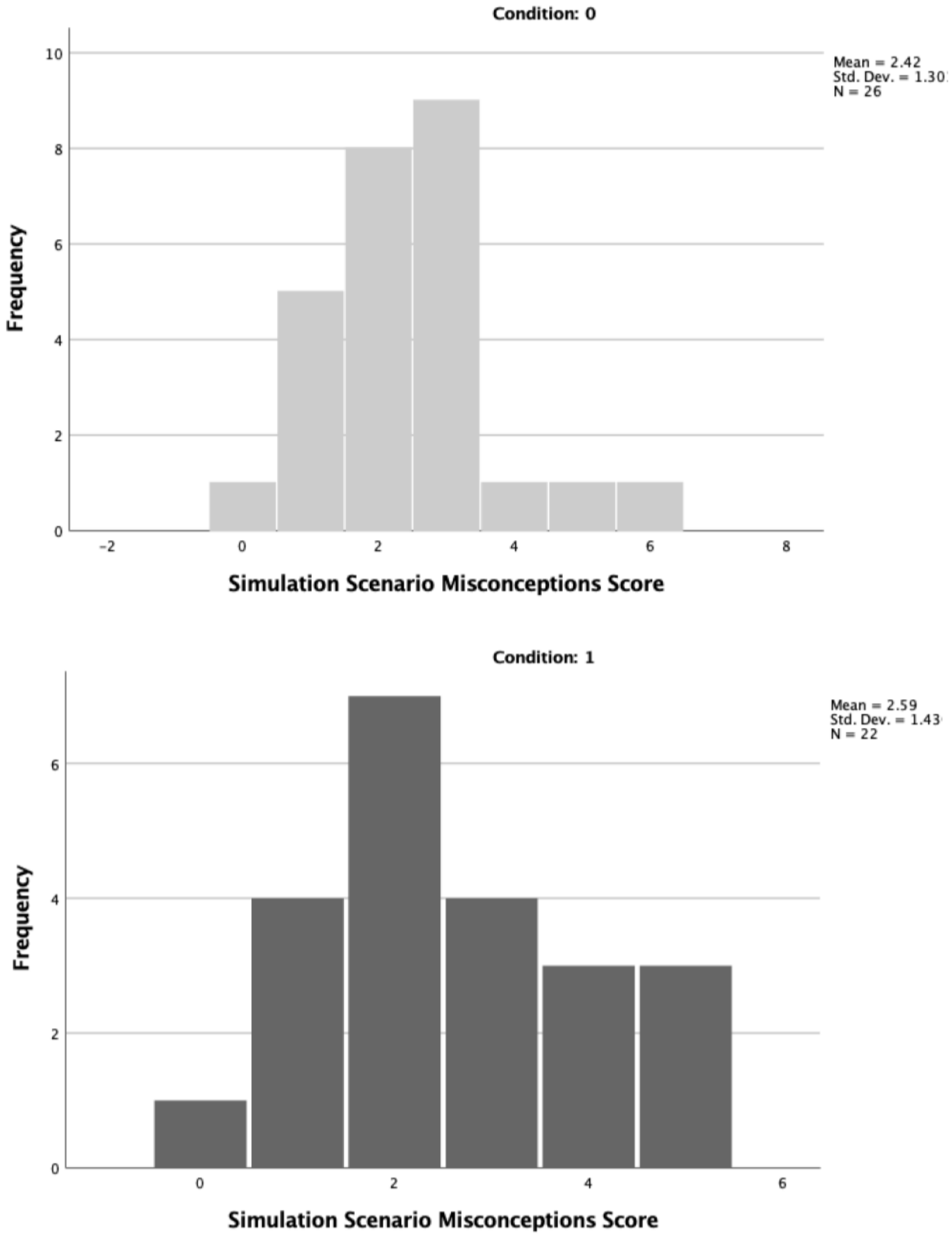
Student	Misconception	Students' Explanations	Score
Student E (Intervention Group)	Misconception 1	Organisms change because changes naturally occur. The DNA within an organism is the cause of any changes that we observe. A mistake or simple mutation in the writing of genes is known as a mutation and can be repeated during reproduction. An organism does not have the ability to control what features will increase its likelihood of survival.	1 + 1 (an extra point was given because students correctly applied Feature 7a - no control agent in the explanation)
	Misconception 3	Organisms do not have any intentional control over the traits they lose and retain. A loss of a trait can happen due to reproduction rates. If a certain trait becomes less popular, the chances of reproducing this trait decrease.	1 + 1 (an extra point was given because students correctly applied Feature 7a - unintentional in the explanation)
	Misconception 6	Natural Selection takes dozens or more of generations to transform populations. In addition, if a trait is still carried there is always a chance it will be reproduced. The transformation does not occur with the goal of survival, the transformation is instead a result of survival.	1+1 (an extra point was given because students correctly applied Feature 7a - not teleological in the explanation)
Student F (Control Group)	Misconception 4	One of Darwin's principles states that there is a differential survival rate (local adaptation) and that certain characteristics	1+1 (an extra point was given because students

	help or impede an animal in a given environment. Mutations also happen at any given moment and are not caused by selection pressure. Even without selection pressure, mutation occurs and changes the survival rate of the animal either in a good or bad way.	correctly applied DP3 - differential survival rate in the explanation)
Misconception 5	One of Darwin's principles states there is random intraspecies variability (individual variation). This means that the animals from a species in a single generation are all different in various ways- size, color, intelligence, perception, memory, child-rearing, and feeding.	1+1 (an extra point was given because students correctly applied DP1 - intraspecies variation in the explanation)
Misconception 6	One of Darwin's principles states that the accumulation of changes happens over many generations. These changes are very small and insignificant over the course of one, two, or even three generations but after MANY generations the change is so significant that even a new species can arise.	1+1 (an extra point was given because students correctly applied DP5 - Accumulation of changes over many generations in the explanation)

Only eight students from the intervention group and three students from the control group applied their understandings of PAIR-C features or the Darwinian Principles in correcting their selected misconceptions. Although 6 students in the intervention group earned a score higher than three points whereas only 3 students in the control group had a score that was above 3 points (See Figure 12) there were no statistically significant differences in their abilities to correct misconceptions measured by this misconception activity ($M_{ctrl} = 2.42$, $SD_{ctrl} = 1.30$; $M_{intv} = 2.59$, $SD_{intv} = 1.44$, $t = .421$, $p = .676$).

Figure 12.

Distribution of Simulation Scenario Misconception Scores



Note. Condition 0: Control group, Condition 1: Intervention group.

Students' Overall Performance on Misconceptions

To understand students' overall performance on misconceptions, I considered both their responses to the pre-posttest open-ended questions and their answers to the misconception explanation activity in the module. Specifically, in analyzing students' overall responses to the pre-posttest questions, I attended to the frequency of each misconception category as well as each PAIR-C feature. In addition, I also paid attention to the distribution of students' self-selected misconceptions in the module misconception explanation activity.

According to Figure 13 and Figure 14, it is evident that students' overall performance on misconception categories improved from the pretest to the posttest. The control group students had a decrease of 14 (from 52 total to 38 total) versus the intervention group students had a decrease of 25 (from 51 total to 26 total). Note that the frequency distribution of different misconception categories was calculated by non-repetitive summing (when a student showed multiple instances of the same misconception category, we only counted once for that specific category of misconception).

Meanwhile, the two figures also show the discrepancy in students' performance in each misconception category. For instance, students in the Intervention group showed more misconceptions about MC2 (Use and Disuse) and MC5 (Anthropomorphism) than the control group on the pretest. However, their performance on the posttest showed fewer frequencies of these two categories of misconceptions.

Figure 13.

Distribution of Misconception Categories in the Pretest

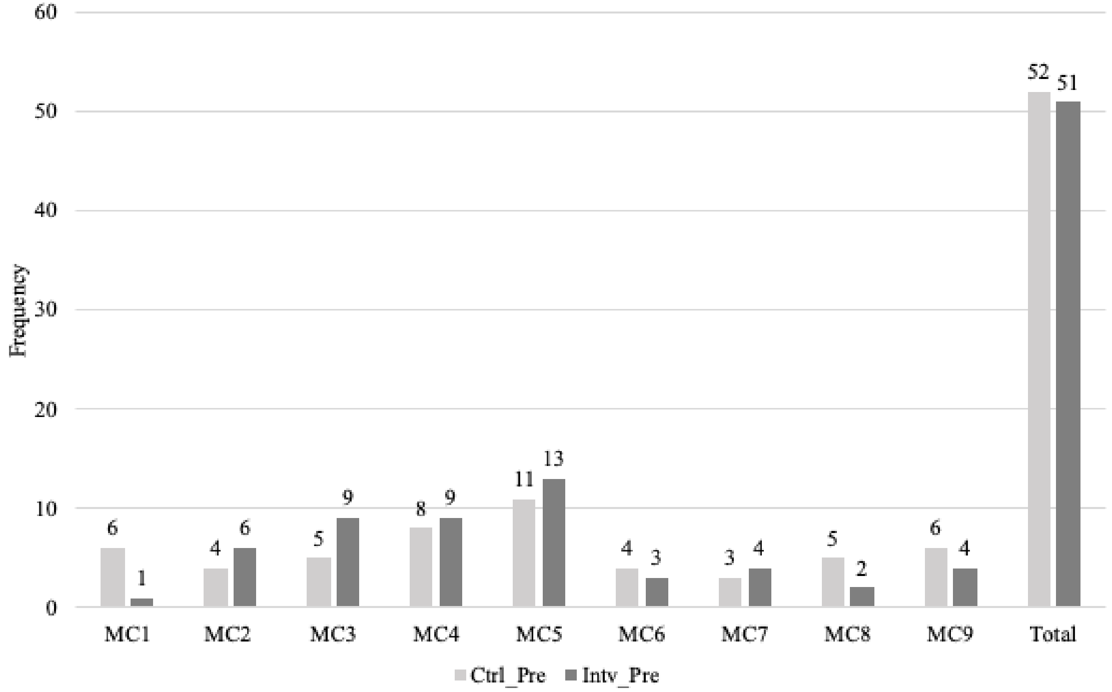
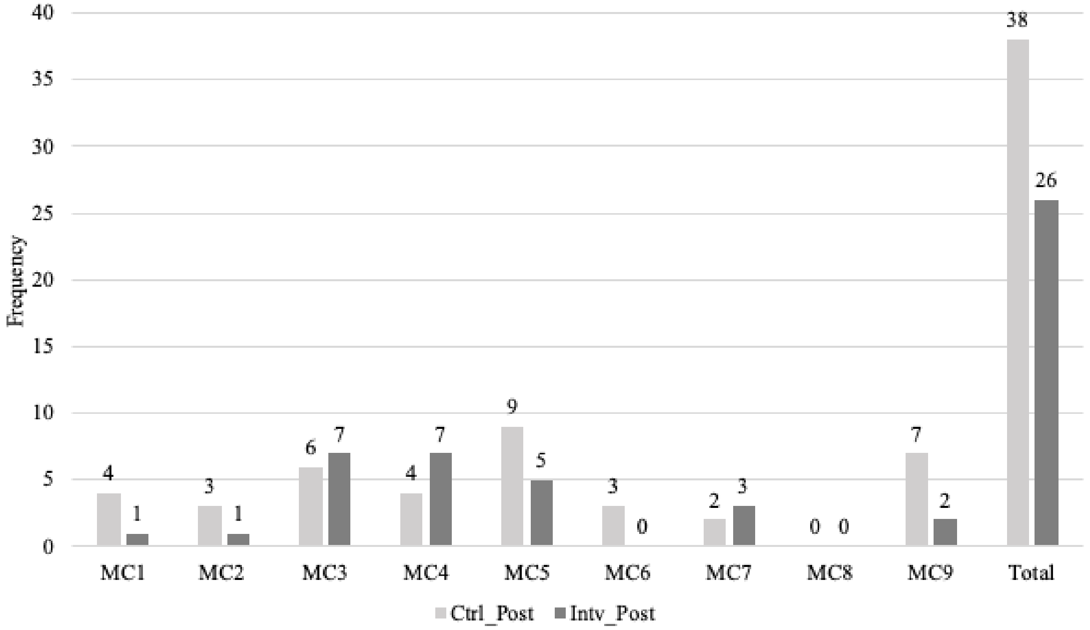


Figure 14.

Distribution of Misconception Categories in the Posttest



Figures 15 and 16 present students' overall performance on PAIR-C features from the pretest to the posttest. Students in both groups demonstrated similar frequencies for the ICR and interaction PAIR-C features on the pretest. However, students in the intervention group had a more drastic decrease for each dimension of PAIR-C features (ICR features decreased from 35 to 20; interaction features decreased from 13 to 3) on the posttest. In contrast, students in the control group did not show such a drastic decrease. Non-repetitive summing was again used in calculating the frequency distribution for each PAIR-C feature.

Figure 15.

Distribution of PAIR-C Features in the Pretest

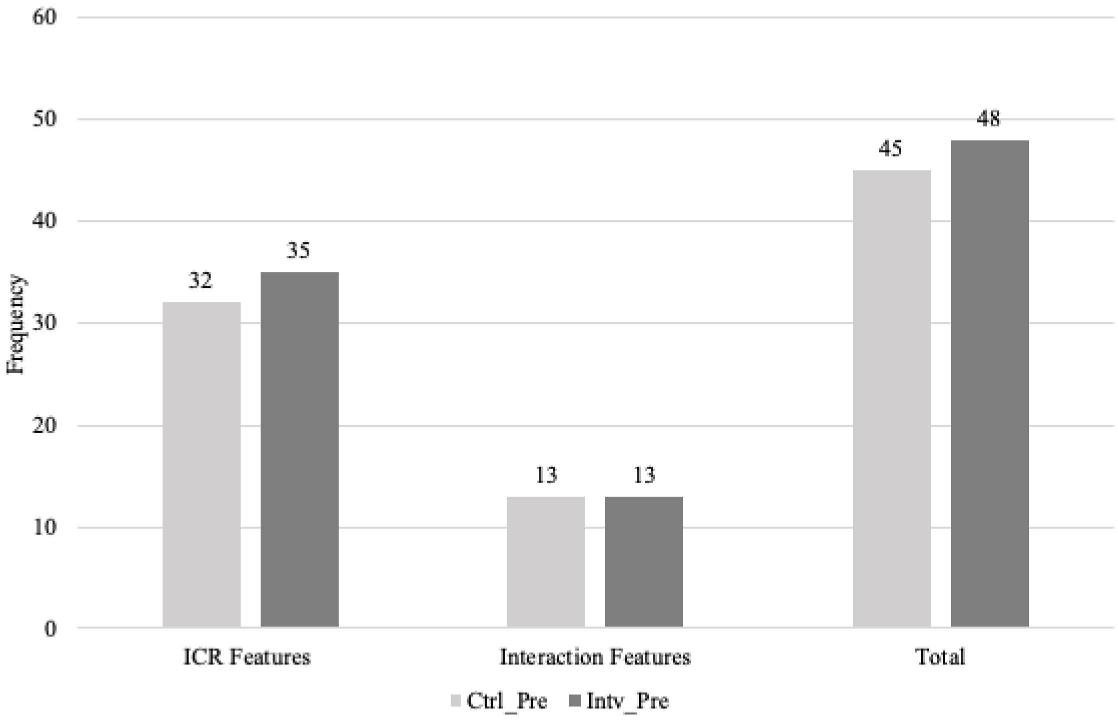
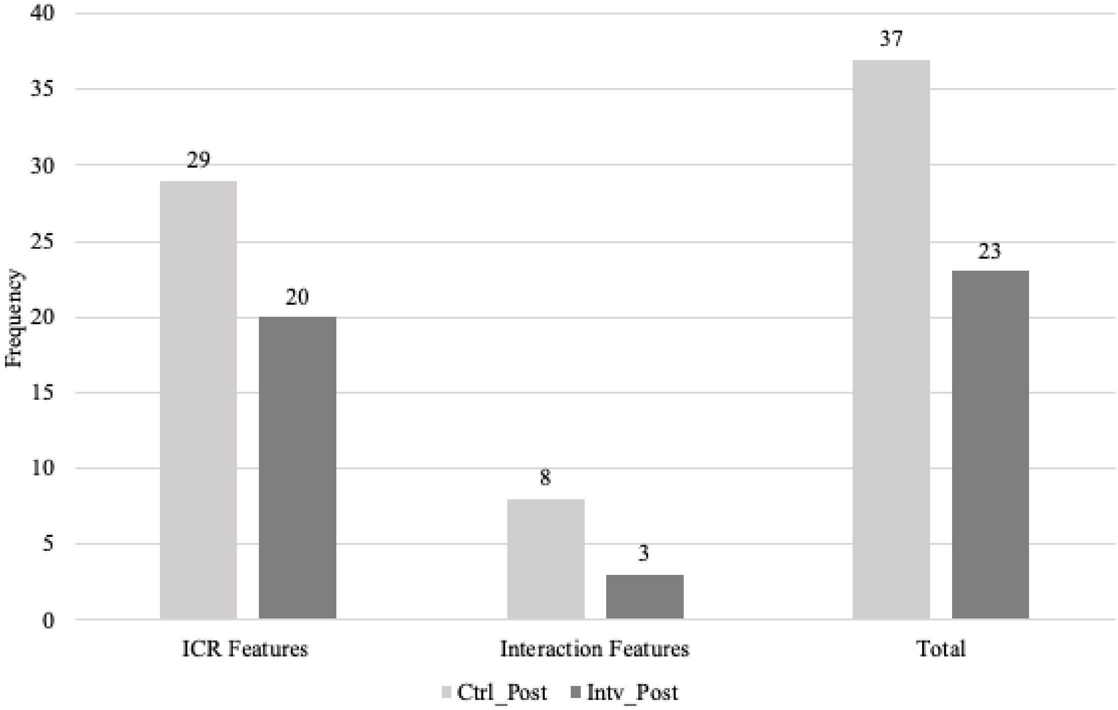


Figure 16.

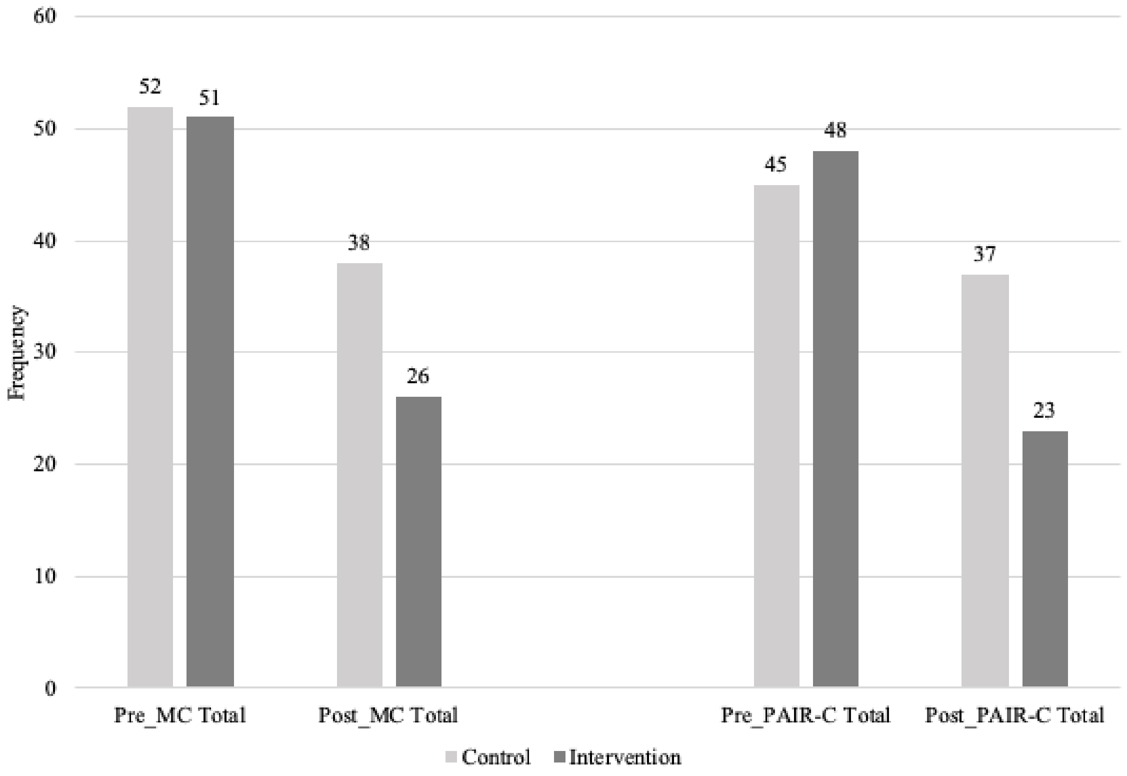
Distribution of PAIR-C Features in the Posttest



To summarize, the two-dimensional coding rubrics of misconceptions showed similar patterns (Figure 16). In the pretest, students of both groups held a similar number of misconceptions (52 vs. 51) which could also be reflected by having a similar amount of PAIR-C features (45 vs. 48). On the contrary, in the posttest, students from the intervention group showed fewer misconceptions and fewer PAIR-C features than the control group although both groups of students had a decrease in both misconception categories and the PAIR-C features. These results suggest that the PAIR-C framework could be translated into a reliable and parsimonious approach when coding students' misconceptions.

Figure 17.

Comparing Distribution of Misconception Categories & PAIR-C Features in the Pre-Posttest Across Groups



In addition to the pre-posttest, students' overall performance on misconceptions was also manifested in their answers to the misconception explanation activity in the module. Because students were most likely to select misconceptions and provide explanations on those items that they were more confident in correcting, the results of this activity could be regarded as a proxy measure for understanding the differences in students' confidence-level in correcting different categories of misconceptions.

Figure 18.

Control Group Percentage Chart for Self-Correcting Misconceptions

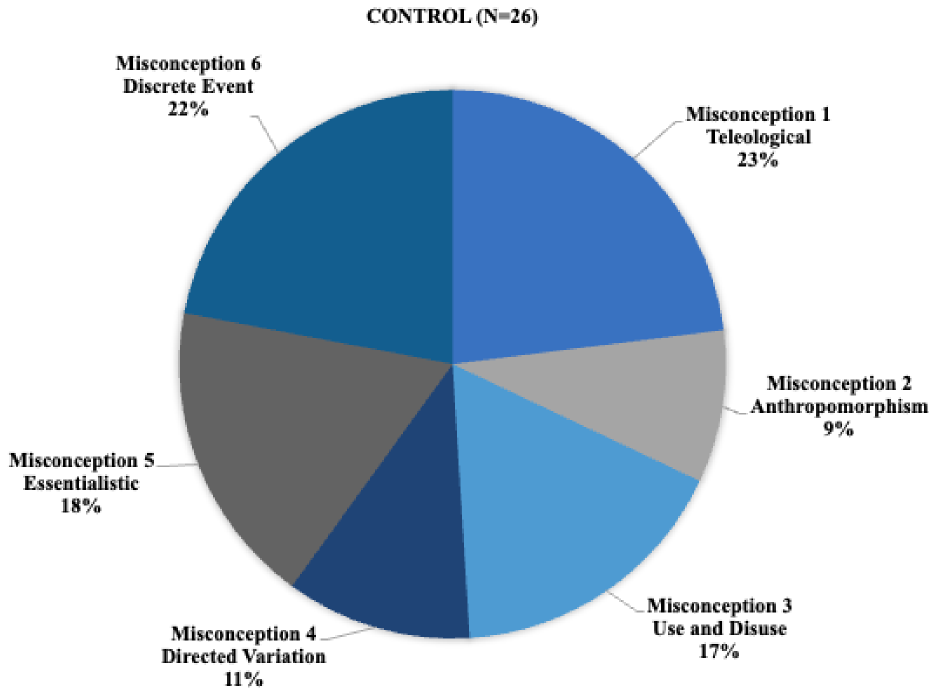


Figure 19.

Intervention Group Percentage Chart for Self-Correcting Misconceptions

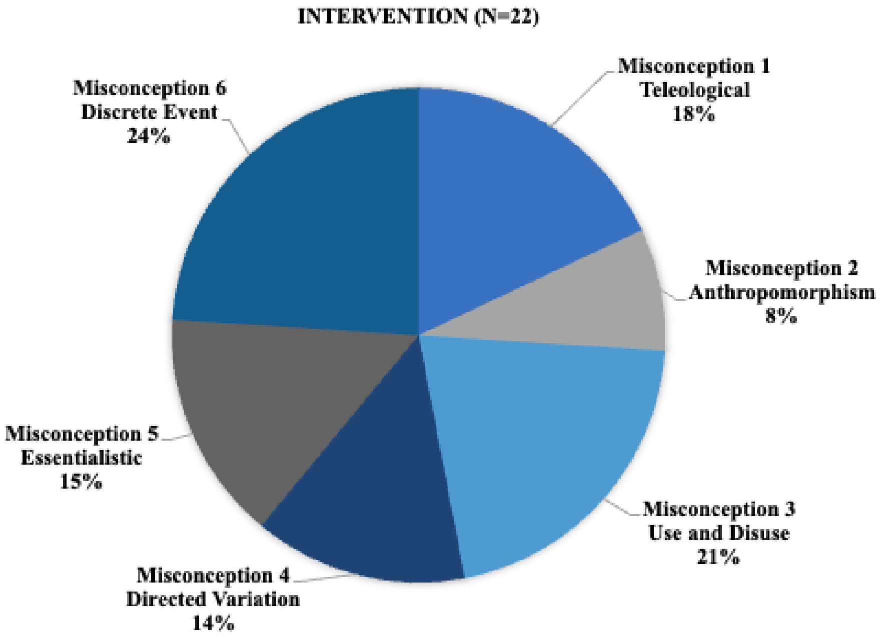


Figure 18 shows that students in the control group selected to correct Misconception 1 (Teleological) and 6 (Discrete Event) more than the other misconceptions. Figure 19 shows that students in the intervention group selected Misconception 3 (Use and Disuse) and 6 (Discrete Event) more than the rest of the misconceptions. Among all the misconceptions, Misconception 2 on anthropomorphism had the lowest probability to be selected from both groups.

This finding echoes what was previously found in students' misconceptions in the pre-posttest. Since a large proportion of students from both groups chose to explain why Misconception 6 (Discrete Event) was incorrect, it indicates that they were confident in providing a correct explanation on viewing natural selection as a process rather than discrete events. This confidence led to a complete eradication of MC8 (events vs. processes) in the posttest (See Figure 14) for both groups. Moreover, as the control group was more confident in explaining Misconception 1, they showed less MC4 (teleological conceptions) than the intervention group in their posttest performance (See Figure 14). Meanwhile, as the intervention group was more confident in explaining Misconception 3, they displayed a drastic drop of MC2 (use and disuse) from 6 to 1 in the posttest, which outperformed the control group (See Figure 14).

Furthermore, to understand the relationship between students' overall performance on misconceptions and their learning outcomes, I conducted correlation analyses. Given that student who correctly understands the inter-level causal relationships when explaining emergent processes of natural selection may hold fewer misconceptions, analyses on the relationship of posttest scores with the misconception variables might provide a source of valid evidence for both the test instrument and the coding results.

Table 16.

Correlations among Posttest, Misconceptions, and PAIR-C Manifestation

	Posttest Results		Misconceptions			PAIR-C Manifestation	
	Total	Deep	Categories	Amount	SelCorrect	ICR	Interaction
Posttest							
Total Score	1.00						
Deep Question	.901**	1.00					
Misconceptions							
Categories	-.461**	-.381**	1.00				
Amount	-.438**	-.392**	.940**	1.00			
Select & Correct	.700**	.609**	-.303*	-.270	1.00		
PAIR-C							
ICR	-.509**	-.422**	.889**	.867**	-.309*	1.00	
Interaction	-.018	-.055	.370**	.448**	.050	.102	1.00

**. Correlation is significant at the 0.01 level (2-tailed)
*. Correlation is significant at the 0.05 level (2-tailed)

As expected, the correlation results showed that students' misconception performances (number of misconceptions categories, amount of misconceptions, abilities to correct misconceptions in the Select & Correct activity) had significantly negative correlations with their posttest performance. Moreover, their PAIR-C manifestations of the sequential ICR features also showed a significant negative correlation with the post-test results. These results indicate that students who performed well on the posttest showed fewer misconceptions, and categories of misconceptions, and used fewer PAIR-C features that manifest individualistic causal structure. Another important piece of information presented in the Table was that no correlation was found between the ICR features and the Interaction features which provided validity evidence of coding the PAIR-C features into two major dimensions to represent independently different constructs.

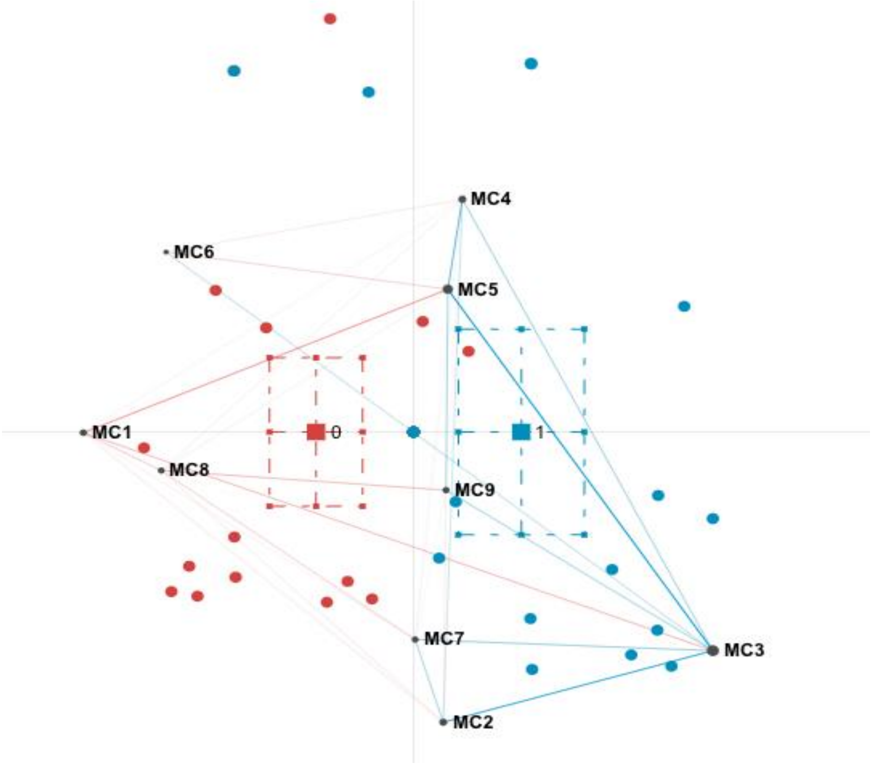
Comparing Students' Misconception Features Using ENA Model

To examine how students' misconceptions change after the modules for both the control and the intervention group, and to compare the differences between students' misconception features between the two groups, Epistemic Network Analysis (ENA) was used for modeling the structure of connections in the misconception data. Specifically, ENA models the connections between Codes by quantifying the co-occurrence of Codes within responses, producing a weighted network of co-occurrences, along with associated visualizations for the data. The thicker edges between the two codes represent a higher frequency of occurrence between these two codes.

Figure 20 visualizes networks of students' misconceptions in the pretest and compares the networks between the control group (coded as 0, shown in red lines) and the intervention group (coded as 1, shown in blue lines). Along the X axis, a two-sample t-test showed the intervention group (mean = 0.28, SD = 0.34, N = 19 was statistically significantly different at the alpha = 0.05 level from the control group (mean = -0.25, SD = 0.26, N = 21; *t* (34.07) = 5.49, *p* = 0.00, Cohen's d=1.76). From Figure 20, it is evident that students in the control group had more frequent co-occurrences of misconceptions between MC1 (Intra-generational change) and MC5 (Anthropomorphism) than the intervention group. In comparison, students in the intervention group showed more co-occurrences of misconceptions between MC3 (Directed Variation) and MC5 (Anthropomorphism), MC2 (Use and Disuse) and MC3 (Directed Variation), as well as MC4 (Teleological beliefs) and MC5 (Anthropomorphism). The comparison between the intervention group (blue) and the control group (red) on the pretest revealed that the intervention group exhibited a higher frequency of misconception co-occurrences. This was supported by the observation that the blue connections showed a more complex structure than those of the control group.

Figure 20.

ENA between Two Groups on Pretest Misconceptions

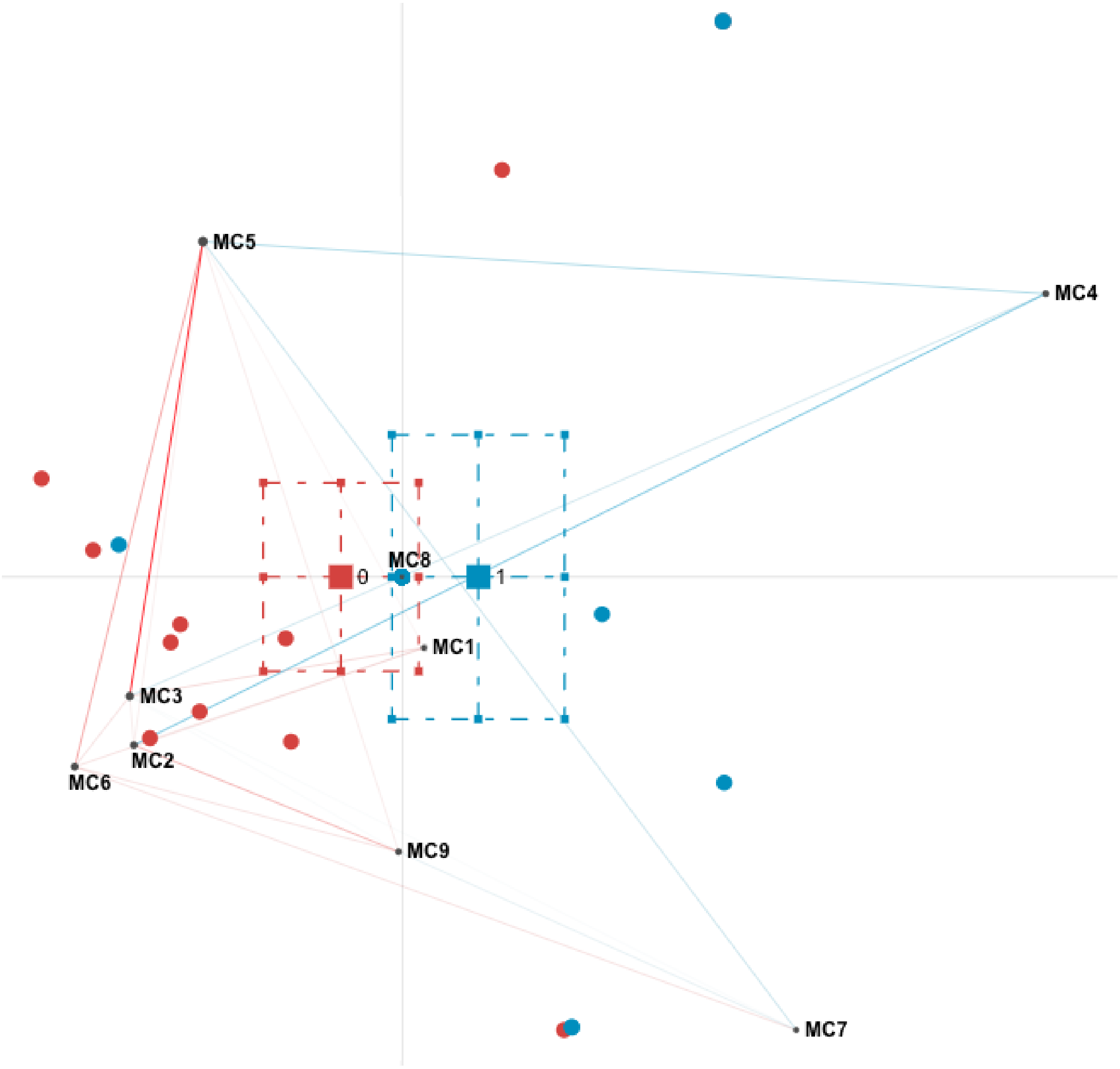


Note. Each blue dot represents a student data point from the intervention group; Each red dot represents a student data point from the control group.

To see how students' misconceptions change after learning the modules, another ENA plot was generated by only examining posttest misconceptions (See Figure 21).

Figure 21.

ENA between Two Groups on Posttest Misconceptions



As shown in Figure 21, there are fewer data points of misconception co-occurrences that happened in the intervention group: the structure of the blue color connection showed a structure with fewer connections between Codes including MC4 and MC5, MC4 and MC2, MC4 and MC8, and MC5 and MC7. In contrast, there are more data points in the control group which indicates more co-occurrences of misconceptions. Among these co-occurrences, MC3 (Directed Variation) and MC5 (Anthropomorphism) appeared to be more dominant for the control group. Along the X axis, a two-sample t-test showed the intervention group (mean = 0.13, SD = 0.29, N = 16 was statistically significantly different at the alpha = 0.05 level from the control group (mean = - 0.11, SD = 0.29, N = 20; *t* (32.59) = 2.50, p = 0.02, Cohen's d = 0.84).

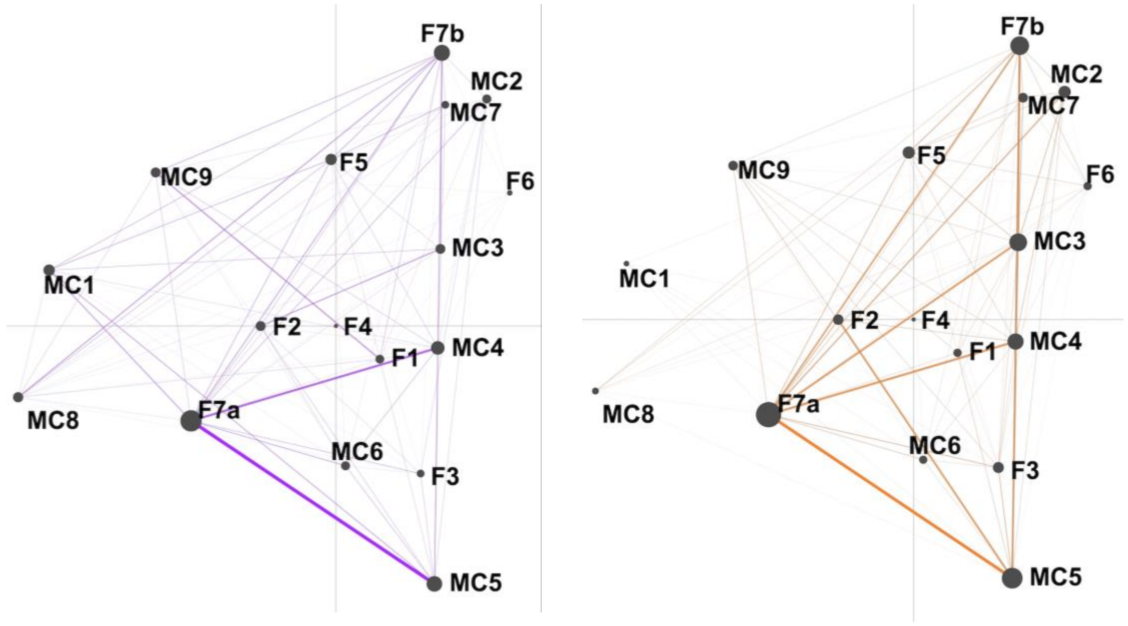
The visualization of the data between the pretest (Figure 20) and the posttest (Figure 21) provided us with lens to see how the modules influence students' misconceptions. In the pretest, the intervention group showed more complicated connections between MC3 and MC2, 4, 5, 6, 7, 9 as opposed to the control group. However, we observed that these connections predominately existed in the control group compared to the intervention group on posttest. The connections in the intervention group shifted to the connections between MC4 and MC2, 5, 8, and MC5 and MC7. It showed that students in the intervention group less often involved misconceptions that they had in the pretest and eliminated them with the support of PAIR-C ABM simulation interventions.

Because we also coded students' misconceptions based on the PAIR-C framework, using ENA could also help visualize the co-occurrences between PAIR-C features and different categories of misconceptions. Figure 22 showed group comparisons on misconception & PAIR-C co-occurrences for the pretest and the posttest respectively. ENA results show the control group network is significantly different from the intervention group in both the pretest and the posttest (p < .001), indicating the co-occurrences between PAIR-C features and students' misconceptions were significantly different in both groups.

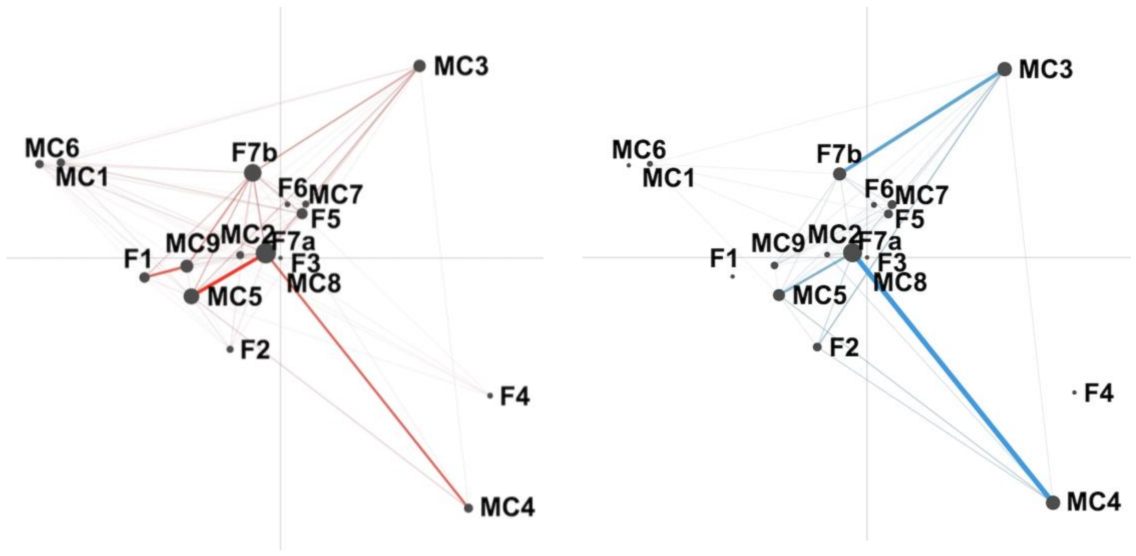
When looking at the following two networks (Figures 22 A & B), F7a (Single causal agent) and MC5 (Anthropomorphism) co-occurred mostly in the pretest for both groups. When looking at the bottom two networks (Figures 22 C & D), it seems that F7a co-occurred frequently with MC4 (Teleological) and MC5 for both groups in the posttest. Additionally, the control group had more co-occurrences between F1 (Distinct set) and MC9 (Absolutes) in the post-test; the intervention group had more co-occurrences between F7b (Align) and MC3 (Directed Variation) in the post-test.

Figure 22.

Comparisons of Two Groups Pre-Posttest Misconceptions & PAIR-C Co-occurrences



A. Pretest Control Group *B. Pretest Intervention Group*



C. Posttest Control Group *D. Posttest Intervention Group*

When comparing the two pretest networks in the first row with the two posttest networks in the second row, it is apparent that most of the nodes in the pretest are distant from each other while the nodes in the posttest are more strongly connected and forming clusters. This can suggest that there is a weaker relationship or dependence between the nodes before students participate in the modules. One reason for this pattern could be that students hold many categories of misconceptions in one statement containing multiple idea units. And when coding PAIR-C features for each idea unit, a co-occurrence relationship between a PAIR-C feature with one misconception cannot be clearly reflected at the whole statement level.

However, after the modules, students in both groups improved their understanding, showed less intertwined misconceptions, and should be able to avoid some categories of misconceptions. Although not all misconceptions were eradicated, the remaining ones were more independent and more clearly related to the specific PAIR-C features. This led to a stronger relationship between the PAIR-C and misconception nodes, and a conceptual cluster of these nodes.

By closely examining the cluster of nodes that appeared in the posttest networks, the PAIR-C framework and its features could be translated into meaningful instructional practices. For example, more explicit instructions on the implication features (F7a, F7b) were needed because three major misconception nodes (MC3, MC4, MC5) were connected to them. This could also inform future revisions of the PAIR-C module. In the current PAIR-C module, I did not explicitly connect the instructions on how to use the two implication features to correct misconceptions such as directed variation, teleological beliefs, and anthropomorphism.

Discussion

In Chapter 7, findings shows that students in the PAIR-C intervention group had a statistically significant reduction of misconception categories and misconception numbers from the pretest to the post-test compared to the Regular control group. However, there were no statistically significant differences between the two groups in students' posttest performance on misconceptions as well as their abilities to correct misconceptions in a module activity.

In the analysis, all the PAIR-C features were incorporated into the codebook to analyze students' misconceptions. Students in both groups demonstrated similar frequencies for the ICR and interaction PAIR-C features on the pretest. However, students in the intervention group had a more drastic decrease for both dimensions of PAIR-C sequential features (ICR features decreased from 35 to 20; interaction features decreased from 13 to 3) on the posttest. In contrast, students in the control group did not show such a drastic decrease.

Correlation results show that students who performed well on the posttest demonstrated fewer categories of misconceptions, and fewer amount of misconceptions, and their misconceptions were less connected with sequential PAIR-C features that manifest individualistic causal structures.

Furthermore, to compare students' misconception features across groups, ENA analytical technique was used. Results show that there were fewer cases of misconception co-occurrences that happened in the intervention group compared to the control group on the posttest. The difference was statistically significant. The ENA method was also used to visualize the co-occurrences between PAIR-C features and different categories of misconceptions. For both groups, nodes in the pre-test were distant from each other while they are more strongly connected and forming clusters in the post-test, suggesting changes in students' misconceptions from an intertwined pattern to a more focused pattern. The remaining co-occurrences of misconception and PAIR-C features provide implications for future instructional design and practices to eradicate common misconceptions more precisely by teaching specific PAIR-C features in specific circumstances.

CHAPTER 8

STUDENT ENGAGEMENT WITH ONLINE MODULE

Introduction

In addition to answering the three main research questions associated with the design of the learning modules, students' deeper understanding, and students' misconception characteristics, this study also aims to understand students' learning processes by examining how they engage differently with the online modules. The Interactive-Constructive-Active-Passive (ICAP) framework (Chi & Wylie, 2014) was used to analyze learner engagement during the modules in an objective manner. Specifically, In Chapter 8, I present the findings on two aspects, including 1) students' cognitive engagement score when responding to prompt questions and how it relates to their learning outcomes, and 2) comparing students' engagement during the learning processes captured by simulation interaction videos.

The ICAP Framework

Chi and Wylie (2014) proposed the ICAP framework, which categorizes the learning process into four distinct modes: passive, active, constructive, and interactive. This framework enables researchers to map learning behavior to these modes and gain insight into how the learner's knowledge evolves. Specifically, passive learning involves learners receiving information from instructional materials without any overt action. Active learning, on the other hand, requires learners to exhibit visible behavior or physical manipulation, such as adjusting simulation parameters to observe changes and pausing or rewinding instructional videos. Constructive learning involves learners generating or producing additional outputs beyond what is provided in the learning materials, such as explaining reasons for the observed changes in the simulation and reflecting on the instructional materials while identifying what was important and missed in their understanding. Interactive learning, not only requires students' utterances to be co-constructive but also needs sufficient turn-taking to occur.

The ICAP framework regards student learning as a gradual process in which students undergo different knowledge-changing stages from passive to active to constructive to interactive and as a result, learning will increase. Because the framework provided descriptions for different modes of learner engagement, it has been widely adopted as an objective measure to understand and evaluate students' learning processes (Hsiao et al., 2022; Sailer et al., 2021).

Findings

Student Cognitive Engagement Score

Because the PAIR-C modules and the Regular modules were designed to be different only in some instructional texts and visual representations. Students in both groups were given equal opportunities to interact with the same simulation models, respond to the same prompt questions, and follow instructions to complete the same length of activities. Therefore, it was assumed that students in both groups should be engaged with the corresponding modules in similar ways.

Results show that there was no statistically significant difference between the two groups in terms of their cognitive engagement score when responding to prompt questions. Specifically, the control group students gained an average engagement score of

89.46 while the intervention group students gained an average engagement score of 87.83 (*t* = -.481, p = .633). Moreover, students in both groups showed an average of approximately 33% constructive engagement, 61% active engagement, and 6% passive engagement when responding to the prompt questions. Therefore, the following analyses did not examine group differences but focused on understanding student overall engagement.

To explore how student cognitive engagement scores relate to learning outcomes, Pearson's correlation was calculated to confirm the relationship between the ICAP percentage in student prompt responses and learning outcomes. Table 17 reveals significantly positive correlations between constructive and the total score ($r = .562$, $p < .001$) as well as the deep question test score ($r = .549$, $p < .001$). There were also significant negative correlations between active and the total score ($r = -.433$, $p < .01$) as well as deep question test score ($r = -.458$, $p < .001$). Significant negative correlations were also found between the passive percentage and the total score ($r = -.365$, $p < .01$) as well as the deep question test score ($r = -.304$, $p < .05$).

To further explore student learning performance, all participants were then divided into two groups according to their gain scores on natural selection. The highest-scoring 50% of learners were allocated to a high-learning performance group and the rest of the learners were assigned to a low-learning performance group. Based on the average gain score of 2.83, 25 learners were assigned to the high-learning performance group and 25 to the low-learning performance group. Then, to determine the differences between the high and low-learning performance groups, an independent *t*-test was conducted, as shown in Table 18. Significant differences in passive ($t = -1.96$, $p = .028$), active ($t = -3.33$, *p*

$< .001$), and constructive ($t = 4.49$, $p < .001$) indicate that learners with higher learning performance on the post-test scores exhibited higher constructive percentage and a lower passive and active percentage than those with lower learning performance.

Additionally, Pearson's correlation coefficient was again calculated to confirm the relationship between the ICAP percentage and the test scores in the high and low learning performance groups respectively. The relationship for the low-performance group is shown in Table 19. Compared with Pearson's correlation for all learners, there is no statistically significant correlation between the test score and constructive ($r = .01$, $p > .05$), between the test score and active ($r = .10$, $p > .05$), and between the test score and passive ($r = -.11$, $p > .05$).

Table 17.

Correlation Coefficient Between ICAP Percentage and Learning Outcomes

	Passive	Active	Constructive	Total Score	Deep Score
Passive	-				
Active	.003	-			
Constructive	$-.557^{***}$	$-.832^{***}$	-		
Total Score	$-.365^{**}$	$-.433^{**}$	$.562^{***}$	-	
Deep Score	$-.304^{*}$	$-.458^{***}$	$.549^{***}$	$.901^{***}$	-

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 18.

Differences Between High and Low Learning Performance Groups

	Group	M	SD	t
Passive	High	5.49%	11.66%	**-1.96****
	Low	11.36%	9.38%	
Active	High	57.03%	13.70%	**-3.33*****
	Low	68.28%	9.84%	
Constructive	High	37.48%	15.31%	**4.49*****
	Low	20.36%	11.38%	

Note. $^{**}p < .01$, $^{***}p < .001$

Table 19.

Relationship Between ICAP% and Test Scores in Low-Performance Group.

	L_Passive	L-Active	L-Constructive	L_Total Score	L_Deep Score
L_P	-				
L_A	-.364	-			
L_C	-.568**	-.560**	-		
L_Total	-.110	.101	.008	-	
L_Deep	-.004	.077	-.064	.767***	-

Note. $^{**}p < .01$, $^{***}p < .001$

The relationships for the high-performance group are shown in Table 20. Compared with Pearson's correlation for all learners, there is no significant correlation between the test score and passive (r = -.10, p > .05). There is a significant negative correlation between the test score and active (r = -.46, p = .037). There is a significant positive correlation between the test score and constructive (r = .443, p = .045**)**.

Table 20.

Relationship Between ICAP% and Test Scores in High-Performance Group.

	H_Passive	H-Active	H-Constructive	H_Total Score	H_Deep Score
H_P	-				
H_A	.252	-			
H_C	-.462*	-.975***	-		
H_Total	-.103	-.457*	.443*	-	
H_Deep	.034	-.497*	.448*	.684***	-

Note. $^{*}p < .05$, $^{***}p < .001$

Based on the ICAP framework, it is hypothesized that students who engage constructively with the learning materials should have higher learning outcomes. To test the ICAP hypothesis, I regressed students' natural selection posttest scores on their percentage of constructive engagement (constructive% was calculated by the number of constructive coded responses divided by the number of all coded responses.) Not surprisingly, the constructive percentage variable explained 67.9% of the variance in students' post-test scores, and the overall regression equation was statistically significant (F [2, 47] = 20.10, p < .001). This finding reconfirmed the ICAP hypothesis and suggests that students who had a higher percentage of constructive engagement during the learning processes would have better learning outcomes. In other words, when responding to the simulation prompt questions, if students put more cognitive effort into providing an explanation, a prediction, or a reflection instead of merely describing, copying information, or manipulating the simulations, they would have better learning outcomes.

Because the modules of the two scenarios were designed to include a total of 15 active prompts and 28 constructive prompts, so ideally, students were expected to have 35% active engagement and 65% constructive engagement. However, the average engagement level for students of both groups were approximately 61% active engagement, 33% constructive engagement, and 6% passive engagement. The relatively low level of cognitive engagement may lead to a reduced effect on the experimental materials.

Comparing Student Engagement During the Learning Processes

Given that the average cognitive engagement level for students when responding to prompt questions was relatively low, I decided to delve deeper into students' overall learning processes in examining how student engagement during the online modules would affect their learning gains. Specifically, I selected four students to do a comparative analysis of their engagement during the first Peppered Moths simulation scenario. The four students were selected based on the following criteria:

- Students successfully video-recorded their interactions with the first module
- Students correctly shared their screens in the interaction video which showed two different screens. One screen was the instructional page with prompt questions and the other screen was the NetLogo simulation interface.
- Two out of the four students were from the Control group and the other two were from the Intervention group

- Within each group, one student had high learning gains in terms of answering deep questions and the other one had low learning gains

Based on the above criteria, four students were selected, and their profiles were listed in Table 21 below. The students' names are pseudonyms.

Table 21.

Student Profile Comparison

Name	Group	Learning Gains	Other Demographic Information
Eric	Intervention	High	Male, 35yrs, other ethnicities Undergraduate Junior, History major Took 3 biology courses in high school Heard about and used NetLogo simulation before Pretest Deep Question score was 1
Justin	Intervention	Low	Male, 19yrs, white Undergraduate Junior, Elementary Education major Took 2 biology courses in high school Never heard about NetLogo before Pretest Deep Question score was 1.5
Alice	Control	High	Female, 21yrs, white Undergraduate Freshman, Early Childhood Education Took 1 biology course in high school Never heard about NetLogo before Pretest Deep Question score was 0

Evelyn	Control	Low	Female, 33yrs, American Indian/Alaska Native Undergraduate Sophomore, Educational Studies Took 2 biology courses in high school Never heard about NetLogo before Pretest Deep Question score was 1

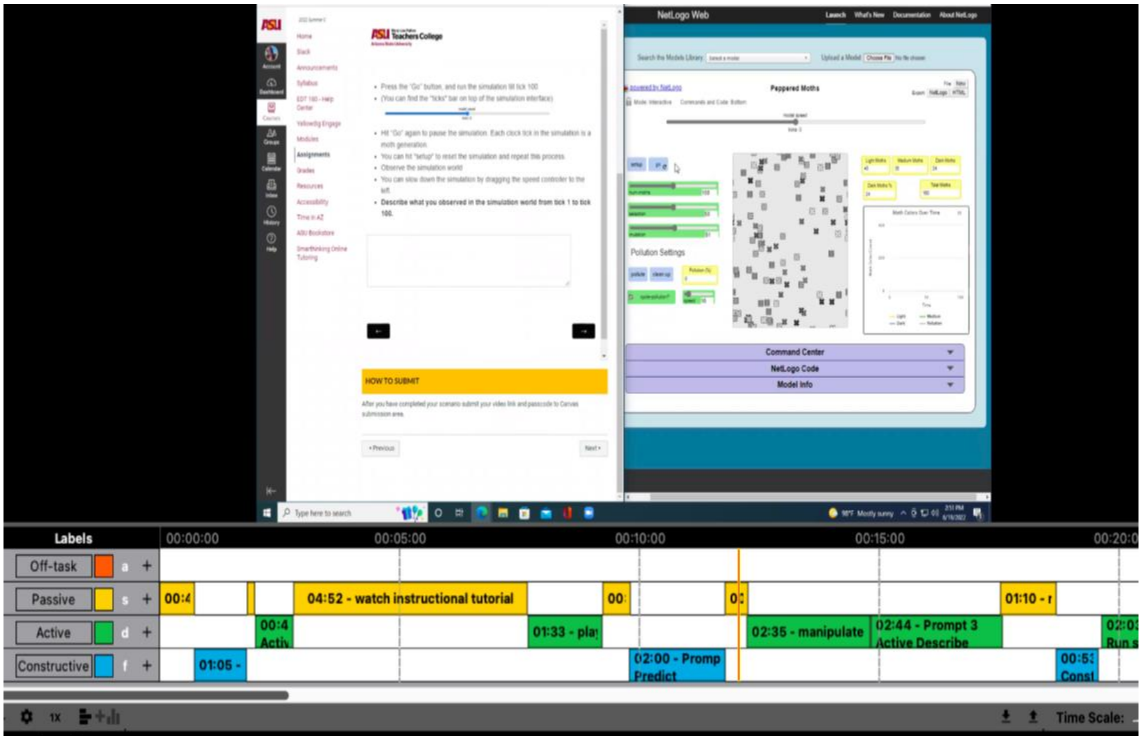
Note. Among the four students, Eric volunteered to participate in the post-project interview and shared more personal experiences with the modules.

Each student's interaction video was imported to the V-note software (Goods, 2019) to conduct coding and analysis based on the ICAP framework. The codebook can be found in Appendix J. Figure 23 shows the coding results for the first 20 minutes of Alice's simulation interaction video. The bottom left side of the figure presents the four color-coded labels/codes: Off-task, Passive, Active, and Constructive. Next to the labels appears the event track. When highlighting a specific learning behavior from the video, I strike a corresponding key on my keyboard to make a notation in the labeling section. Once the notation begins, a colored bar populates the event track from the duration of the event until the key is struck again. For example, the yellow bar that appeared at "00:05:00" represents an active learning event when Alice is watching instructional tutorials.

Based on the coding event intervals, the V-note software automatically generated histograms for each label as shown in Appendix K.

Figure 23.

Visualization of V-note Coding Interface for Alice



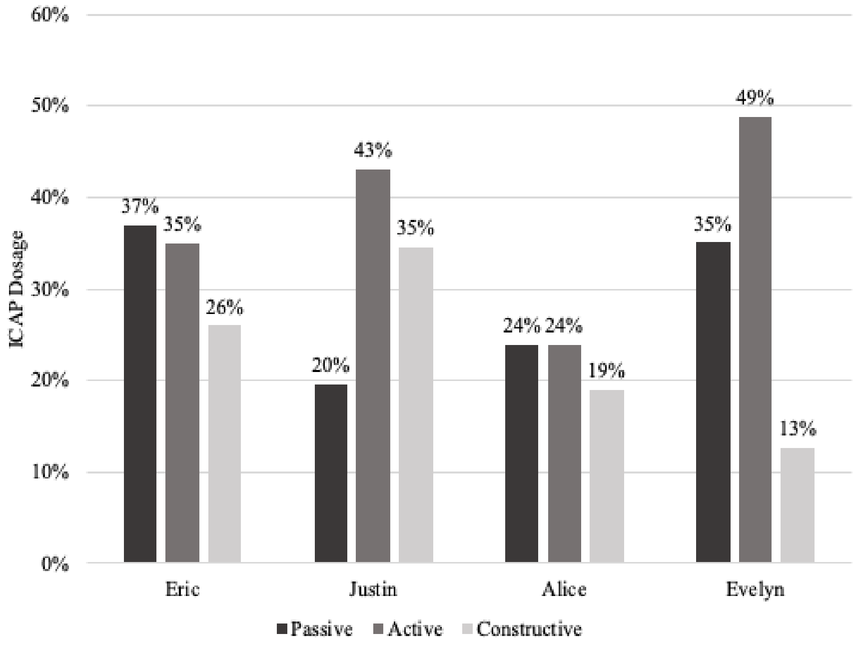
To compare and contrast the similarities and differences between the four students, Figure 24 depicts each student's ICAP dosage. Eric and Alice both had similar dosages for passive and active engagement. In contrast, Justin and Evelyn showed great ICAP dosage discrepancies between the passive and active engagement respectively. Both Justin and Evelyn spent almost half of their learning time conducting active engagement activities (i.e., manipulating simulations and describing changes).

Take Evelyn as an example, in Activity 3, she spent 7 minutes adjusting simulation model parameters to observe pattern changes. In watching her simulation interaction video, it was not hard to find the reason why she spent so much time merely manipulating the simulation. It seemed that she tended to randomly adjust multiple simulation parameters (i.e., selection pressure, mutation, pollution level) at the same time. Therefore, when she hit “run”, the simulation would present results based on the combined effect of all her model adjustments. However, she was supposed to follow the instructions by only manipulating one parameter at a time.

Another finding was that both Eric and Alice spent considerable time reading the instructional text. For example, before answering a prompt question, Eric would go back to the previous instructional page and re-read the text on PAIR-C features. Similarly, when watching a simulation video provided on the instructional page, Alice would pause the video and re-read the above text again by moving her mouse to the upper side of the screen. In contrast, Justin seemed to skip reading the instructional text about PAIR-C and provide explanations or reflections without incorporating ideas from the instructional text. When instructional text and simulation video were presented together on one instructional page, he would only spend time watching the video and skipped the texts.

Figure 24.

Student ICAP Dosage Comparison Chart



These findings suggest that students who had high learning gains (i.e., Eric & Alice) might not necessarily possess the highest dosage for constructive engagement in a self-directed online simulation learning environment. However, they spent a fair amount of time reading the instructional text (passive engagement) accompanied by purposeful simulation manipulations (active engagement) before constructing an explanatory or reflective response (constructive engagement). For students who had low learning gains, one reason for the lack of improvement might be because these students might need more scaffolding or tacit monitoring in a self-directed simulation learning environment. Extensive active manipulation of simulation parameters without clear purposes would hinder students' learning outcomes and might also lead to fatigue and cognitive overload during the online module learning process.

Discussion

In this chapter, I explored how student cognitive engagement relates to the learning outcomes of the online modules. Table 17 shows significantly positive correlations between the constructive percentage and total score, as well as the deep question score. It also shows significantly negative correlations between the passive percentage and the total/deep question scores, and the active percentage and the total/deep question scores. This suggests that responding constructively to the prompt questions is crucial to the higher learning effectiveness of the online modules, whereas passive and active responses are less beneficial. This validates the ICAP hypotheses (Chi & Wylie, 2014) that learning outcomes improve progressively from passive to active, and from active to constructive mode. Accordingly, learners who exhibit a higher constructive percentage in responding to the prompt questions result in higher post-test scores. Conversely, learners with a higher passive percentage in responding to the prompt questions result in lower post-test scores.

Table 18 shows that passive in the high learning performance group is significantly lower than that in the low learning performance group, and constructive in the high learning performance group is significantly higher than those in the low learning performance group. When learning the online module, better learning performance is achieved when learners constructively engage in answering the prompt questions by predicting, explaining, reflecting, and elaborating instead of just describing the observations or simply paraphrasing the instructional texts.

Table 20 presents that only the constructive percentage exhibits a significantly positive correlation with the deep question score in the high learning performance group whereas the active percentage exhibits a significantly negative correlation with the deep question score in the high learning performance group. This indicates that actively responding to the prompt questions by describing the changes observed in the simulation or copy-pasting the instructional texts into the prompt response box is not sufficient to promote effective learning outcomes or deeper understanding. In contrast, learners who constructively cultivate their responses by providing explanations and elaborations can further improve their learning performance.

Since the average cognitive engagement level for students when responding to prompt questions was relatively lower than expected, a more comprehensive examination of students’ overall learning processes was conducted by analyzing four students’ module interaction videos and comparing their engagement behaviors and patterns. These four students were representatives of students from both experimental groups with either high learning gains or low learning gains.

By comparing their ICAP engagement histograms and their ICAP dosages, these four students showed the following patterns in their learning processes which could inform the future design of the online learning module:

1. Students with lower learning gains tended to skip reading the instructional texts versus students with higher learning gains tended to read and refer back to the instructional texts. —> Future design may consider presenting the instructional texts with the audio or within videos which provide students with multi-channel inputs for processing important instructional content.
2. Students with lower learning gains tended to show extensive manipulation of the simulation model parameters versus students with higher learning gains tended to mindfully adjust simulation parameters to test hypotheses. —> Future design may consider limiting students' control over the simulation model parameters or sending reminders to students if they showed signs of excessive manipulation.
3. To avoid fatigue and cognitive overload, future design may consider separating the current 60-minute module (covering one simulation scenario with four activities) into two mini-modules.

CHAPTER 9

IMPLICATIONS AND CONCLUSION

The present study investigates the effectiveness of using the PAIR-C framework to modify agent-based models (ABMs) in fostering students' deep understanding of emergent process concepts, specifically Natural Selection (N.S.). While ABMs have been recognized as useful tools for understanding emergent process concepts, the path toward deep understanding requires careful design. Using ABMs without explicitly elaborating on agents-pattern or inter-level causal relationships may not result in a thorough understanding of emergent process concepts.

Overall, this study contributes to research practices that integrate agent-based models into science curricula by applying the ontological PAIR-C framework to the design of learning modules. Using the MERs-complemented ABM approach, the PAIR-C features were translated into effective online simulation-based module as it teaches natural selection from a system-thinking perspective compared to the traditional Darwinian perspective.

This study also has theoretical and practical implications. Theoretically, the study proposed that the PAIR-C framework could explain why misconceptions are robust and persistent. Meanwhile, the study also proposed that applying the PAIR-C framework into instruction could address robust misconceptions by shifting students conceptual knowledge from individualistic causal structures to collective casual structures. Given that the PAIR-C ontological framework provides two coherent underlying causal structures with opposing features for discriminating between the emergent processes and the sequential processes, grasping the PAIR-C features of emergent processes should help students correctly describe and explain emergent processes without misconceptions.

Practically, this study contributes to the design of online modules that foster deep understanding of complex concept. Using natural selection as the focal concept, students in the PAIR-C intervention group demonstrated better performance in answering the deep questions compared to those in the Regular control group. It remains to be explored whether this finding would be replicated in other domains. Specifically, the findings that learning with a PAIR-C perspective can foster deeper understanding of inter-level causal relationship or correct explanation of the causal mechanisms for the perceptual pattern. Moreover, the approach of integrating the PAIR-C features using MERs-complemented ABM seems promising to be operationalized broadly.

Lessons Learned from the Pilot Studies

Building upon the two online pilot studies, this dissertation project conducted a stand-alone study to compare the effectiveness of the PAIR-C module with the Regular module. Both modules presented students with two different ABM simulation scenarios (i.e., the Peppered Moths and the Rock Pocket Mice) to teach the concept of N.S. In contrast to Pilot study 1, explicit instruction and timely feedback of possible answers were provided to students that demonstrate how the simulation manifested the PAIR-C features and how to use the PAIR-C features to explain the N.S. process, especially how certain pattern transforms from the agent level and the causal mechanisms underlying the observed changes. Although the instruction of the modules used in the current study followed a prescribed structure as in the pilot, it allowed students to interact and explore the simulation themselves rather than merely watching prescribed simulation videos. The current study combined the use of ABM simulations with multiple external representations (e.g., recorded videos and ABM screenshots) to create an interactive and guided learning environment for students. The pilot studies also inform the current study to include questions of different formats in the pre-posttests and simulation prompts, especially questions that are open-ended and constructive to detect students’ differences in their causal reasoning as well as their level of cognitive engagement.

The pilot study was conducted with a small group of elite high school students who had high prior knowledge of N.S. However, a considerable number of students failed to complete the low-stake pre-posttest in a timely manner while participating in the unmonitored online experiment, leading to insignificant experimental results. To address this, the current study recruited a group of college students who on average had taken 1-2 biology courses at high school and were enrolled in an online technology literacy class offered at the School of Education. Most of the participants majored in Education and were pre-service teachers who held an interest in integrating simulations into their classroom instruction. Compared to the participants in the pilot studies, students in the current study were more motivated to learn and were more responsible for their test-taking performance even though they were told that they were not being graded based on the test scores.

Lessons Learned from Designing PAIR-C Module

In comparison with the Regular module which provided direct instruction on the five Darwinian principles, the PAIR-C module provided direct instruction on the PAIR-C features and how to apply these features to explain the emergent properties of N.S. shown in the simulation. Because Chi (2012b) suggested that the four perceivable *Interaction features* were prerequisites for understanding *Inter-level features*, instructions about the four *Interaction features* were given to students prior to introducing the *Inter-level features* in the module. Specifically, students in the PAIR-C group were given multiple external representations (MERs) generated from modified ABMs which used visual cues (e.g., blue links and red links) to represent different types of interactions between agents (e.g., reproduction interaction and predation interaction). Instruction about the four perceivable *Interaction features (i.e., same set, random, simultaneous, and independent)* were given to students while presenting the link visible MERs in the forms of videos, screenshots, and images. Following the instruction on the *Interaction features*, students in the PAIR-C intervention group were then given multiple ABMs screenshots that represent multiple generations to learn about the inter-level feature of *converging change*. In addition, students were also given a link-visible ABM video that recorded pattern change under different mutation levels to introduce the other inter-level feature of *collective summing* causal mechanisms. The rest of the PAIR-C implication features (*no causal agent* and *non-align* between the agents and the pattern) were introduced to students through a data practice activity using the interactive Peppered Moths simulation to collect data and observe the changes in the percentage of dark-colored moths over time.

Results show that students' performance in responding to simulation prompt questions targeting the understanding of the *Interaction Features* had a significant positive correlation with their performance on the post-test ($r = .484$, $p = .022$), the posttest deep questions regarding inter-level causal relationships (*r* = .490, *p* = .020), and a marginally significant positive correlation with their performance on other simulation prompt questions targeting the understanding of the *Inter-level Features* (*r* = .419, *p* = .053). Moreover, students' performance in responding to prompt questions about inter-level features also had a significant positive correlation with their performance on the post-test (*r* = .556, *p* = .007) as well as post-test deep questions (*r* = .655, *p* < .001). These findings inform that teaching the four perceivable PAIR-C interaction features is fundamental to teaching the PAIR-C inter-level features. Students' performance in learning both the *Interaction features* and the *Inter-level features* can influence their post-test performance, especially questions that assessed their deeper understanding of the inter-level causal relationships. Therefore, future designs of the PAIR-C module should consider following the path of teaching *Interaction features* first, and then explicitly elaborating on the *Inter-level features.* Considering a student's prompt response "The PAIR-C features seem well organized and thought out, but it is difficult to memorize them in such a short time frame. I understand the concept of converging change and other elements of PAIR-C but am unable to tell you which ones mean what, unfortunately", more efforts should contribute to instructing the *Inter-level features* by designing activities that are targeted at introducing one *Inter-level feature* at a time. In the current Peppered Moths PAIR-C module, instruction on *Interaction Features 1-4* appeared three times during Activity 2 and 3. However, *Feature 5* was only introduced once at the end of Activity 3. *Feature 6, 7a, and 7b* were introduced in the single Activity 4. Similarly, in the current Rock Pocket Mice PAIR-C module, the four *Interaction features* were instructed or mentioned three times during Activity 2 and 3. In contrast, the *Inter-level features* were only introduced once in the single Activity 3. By allocating more instructional time and designing targeted activities for *Inter-level features* during the module, it might be possible to further improve students' learning outcomes, particularly in fostering a deeper understanding of inter-level causal relationships. Since students using the Regular modules also had significant learning gains from the pretest to the posttest, future efforts should consider embedding the instruction of PAIR-C features into explaining the Darwinian Principles which might lead to more promising results.

Lessons Learned from Investigating Students Deeper Understanding

Chapter 6 investigates the effects of the two approaches to teaching N.S. using agent-based simulations. Results show that the PAIR-C ABM approach is more effective in facilitating students' deeper understanding of N.S. in terms of answering questions about inter-level causal relationships compared to the Regular ABM approach. When asked how students would explain N.S. differently in the post-project interview, students from the PAIR-C ABM intervention group demonstrated a system thinking perspective by explicitly referring to N.S. as a "whole concept" with "multiple factors involved in the process". In contrast, students from the Regular ABM control group did not show such system-thinking expressions. Instead, they focused more on reasoning how the environmental factors could influence N.S.

Nevertheless, when examining students' abilities to explain the process of N.S., students in both groups did not show significant improvement or differences in their responses. Since we identified the reason was due to the lack of instruction on how to apply the PAIR-C features or Darwinian Principles in the elaboration of the N.S. process, future efforts should deliberately teach students about the application and provide more scaffolding when they first explain the process of N.S. Revised instructions were proposed in Table 10 which presented how the PAIR-C features could be used to further elaborate the five-step process of N.S.

Moreover, students in both groups did not show statistically significant improvement nor differences in their performance on answering the far-transfer questions even though students in the PAIR-C intervention group showed a drastic decrease in misconceptions. When asked to reflect on whether there were common properties between the emergent process of natural selection and other emergent phenomena (e.g., geese flying into V shape, fake news spreading) in the post-project interview, interviewees from both groups could not make an explicit connection that used what they learned in the N.S. module to explain other emergent phenomena. The two interviewees from the PAIR-C intervention group did not mention or refer to any PAIR-C feature when responding to this question. Given that no instruction was given to students on connecting the properties of the emergent N.S. process with other emergent phenomena, future efforts should make this connection more explicit to students in designing the module.

Lessons Learned from Characterizing Student Misconceptions

Chapter 7 examines student misconceptions from multiple perspectives and data sources to get a comprehensive understanding of student misconceptions, capture the characteristics of misconceptions, and compare the differences between student misconceptions across groups. Findings reveal that students in the PAIR-C ABM intervention group demonstrated drastically fewer categories and amount of misconceptions towards N.S. in the pre-posttest while those in the Regular ABM control group did not show such drastic changes. Similarly, students in the PAIR-C ABM intervention group also demonstrated drastically fewer PAIR-C features from the individualistic causal structure in explaining N.S. in the pre-posttest while those in the control group did not show a drastic decrease. Nevertheless, there were no statistically significant differences between the two groups in showing misconceptions and PAIR-C sequential features in the posttest as well as their abilities in correcting misconceptions in one of the module activities. There needs to be a more fine-grained analysis to capture and compare the differences between student misconceptions across groups.

Since students' misconception performances were significantly correlated with their post-test performance, it is critical to address common misconceptions explicitly and precisely in the instruction to further improve students' learning outcomes. Results from the Epistemic Network Analysis (ENA) provided some possible solutions for achieving this goal. Specifically, because MC3 (Directed Variation), MC4 (Teleological Conceptions), and MC5 (Anthropomorphism) constantly appeared to be the nodes that strongly connected to other misconceptions in the pre-posttest, more explicit instruction on how to alleviate these common misconceptions was needed in the instruction. Because the ENA results also show that the three major misconception nodes (i.e., MC3, MC4, and MC5) had a strong co-occurrence relationship with two PAIR-C implication features (i.e., Feature 7a: single causal agent, Feature 7b: alignment between agent and pattern), future instruction could consider applying the two PAIR-C implication features to elaborate on those misconceptions and provide correct explanations.

Lessons Learned from Analyzing Student Engagement

Chapter 8 adopts the ICAP framework to assess student cognitive engagement in responding to simulation prompt questions and the overall engagement with the module. Understanding student engagement during the learning process is crucial for predicting their learning outcomes. In line with the ICAP hypothesis, students who showed more *Constructive* engagement in answering simulation prompts tended to have higher post-test scores. In contrast, students who showed more *Passive* engagement in answering simulation prompt questions tended to have lower post-test scores. One thing to be noted is that when students showed more *Active* engagement in responding to constructive simulation prompts (i.e., describing the pattern instead of explaining why it happens, paraphrasing important information instead of reflecting on why the information is important), they were less likely to have high scores in answering post-test questions, especially deep questions about inter-level causal relationships.

The negative effect of *Active* engagement was also found in examining student interaction behaviors with the integrated simulation learning environment. From the comparative analysis of four students' ICAP engagement dosages, it seems that for students who demonstrated the *"Passive < Active > Constructive"* pattern and spent significant amounts of time conducting *Active* engagement activities (i.e., manipulating simulation parameters, observing the changes, describing the observations, etc.) would achieve less in their learning and understanding. However, students who demonstrated the *"Passive = Active > Constructive"* or the *"Passive > Active > Constructive"* pattern would have higher learning gains because spending considerable time reading the instructional texts (*Passive* engagement) as well as pausing or rewinding the instructional videos (*P*assive or *Active* engagement) would prepare students before they constructively write a well-thought explanatory or meaningful reflective response. Apparently, active manipulations of simulation parameters without clear learning purposes or hypotheses in mind would hinder students' learning outcomes.

Therefore, when evaluating students' learning processes, it is important to understand and analyze the distribution of ICAP dosage. It is not always better to promote a higher dosage of *Active or Constructive* engagement in an integrated simulation learning environment. For students who had disproportionately lower *Passive* dosage in the learning process, even having higher Active and Constructive engagement could not lead to high learning outcomes. Future efforts should focus on designing learning modules that could engage students in reading important instructional texts while avoiding excessive manipulation of simulation parameters.

Limitations and Future Directions

This study has several limitations. First, this was a short-term study. Testing after a longer time interval may have shown possible further benefits and differences between the PAIR-C ABM intervention group and the Regular ABM control group. Second, only five students volunteered and participated in the post-project interview. Given that these volunteers received relatively higher scores in the post-test, their responses to the interview questions could not represent low-performing students. Having both high and low performers in the interview may help researchers better understand students' learning of natural selection, their interactions with the online module, as well as their self-efficacy in explaining other emergent phenomena. Third, the modules did not connect the learning of N.S. with understanding other emergent phenomena. Making connections between the properties of N.S. process with the properties of other emergent phenomena during instruction may further help students develop deep understanding of inter-level causal relationships and system thinking for understanding other emergent phenomena and achieve far transfer in their learning. The construction of systems knowledge did not seem to be a “free” byproduct of learning disciplinary emergent process concepts, it needs deliberate instruction and learning design (Samon and Levy, 2020).

Since systems thinking is an important skill for understanding of systems in various disciplines, it is important that teachers know how to assess students’ deep understanding to support reasoning about complex systems and develop theory-based strategies that will advance the instructional practices in teaching emergent process concepts from system-thinking perspective. Therefore, the PAIR-C module designed for this dissertation study can be extended to a teacher (educator) professional development (PD) program which translate the PAIR-C framework into a theory-driven pedagogy that could explicitly show teachers how system thinking can be integrated into science instruction. To effectively integrate systems thinking into science education, it requires collaboration between science teachers, teacher educators, and researchers.

In addition to developing a PD program, I hope to continue this line of work by investigating the effectiveness of PAIR-C module when students are working in pairs. Given that the *Interactive* engagement is the highest level of cognitive engagement (Chi et al., 2018), I am interested in investigating whether the effect of the PAIR-C module could be moderated by student collaboration quality.

Finally, I plan to further revise the PAIR-C module by including more instruction and activities on elaborating the inter-level features and implication features. I also plan to seek long-term partnership with a university biology teacher and an instructional designer to co-create an online module that integrates PAIR-C features into curriculum instruction. The goal is to systematically distribute the PAIR-C module to more college students and further investigate its effects on fostering deep understanding of emergent science concepts as well as emergent phenomena.

REFERENCES

Abraham, J. K., Meir, E., Perry, J., Herron, J. C., Maruca, S., & Stal, D. (2008). Addressing Undergraduate Student Misconceptions about Natural Selection with an Interactive Simulated Laboratory. *Evolution: Education and Outreach, 2*(3), 393–404. https://doi.org/10.1007/s12052-009-0142-3

American Educational Research Association. (2018). *Standards for educational and psychological testing.* American Educational Research Association.

Basu, S., Sengupta, P. & Biswas, G. (2015). A Scaffolding Framework to Support Learning of Emergent Phenomena Using Multi-Agent-Based Simulation Environments. *Res Sci Educ 45,* 293–324. https://doi.org/10.1007/s11165-014-9424-z

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, *27*(5), 415-427 https://doi.org/10.1002/tea.3660270503

Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: A constructionist learning environment for materials science using agent-based modeling. *International Journal of Computers for Mathematical Learning, 14*(2), 81-119. https://doi.org/10.1007/s10758-009-9148-8

Bray Speth, E., Long, T. M., Pennock, R. T., & Ebert-May, D. (2008). Using Avida-ED for Teaching and Learning about Evolution in Undergraduate Introductory Biology Courses. *Evolution: Education and Outreach*, *2*(3), 415–428. https://doi.org/10.1007/s12052-009-0154-z

Centola, D., Wilensky, U., & McKenzie, E. (2000). Survival of the groupiest: Facilitating students' understanding of the multiple levels of fitness through multi-agent modeling - The EACH Project. *International Journal of Complex Systems*, 12–15.

Chi, M. T. H., Su, M., VanBibber B., Cho, J., Ding L., Boucher, N. (Under Review). *Individualistic versus collective causal structures.*

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, *6*(3), 271-315. https://doi.org/10.1207/s15327809jls0603_1

Chi, M. T. H. (2005). Common sense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*, 161–199. https://doi.org/10.1207/s15327809jls1402_1

Chi, M. T. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes. In *International handbook of research on conceptual change* (pp. 49-70).

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science, 18*(3), 439-477. https://doi.org/10.1016/0364-0213(94)90016-7

Chi, M. T., Kristensen, A. K., & Roscoe, R. (2012a). Misunderstanding emergent causal mechanism in natural selection. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 145-173. New York, NY: Oxford University Press.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chi, M. T., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., ... & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive science, 42*(6), 1777-1832. https://doi.org/10.1111/cogs.12626

Chi, M. T., & Roscoe, R. D. (2002). The processes and challenges of conceptual change. *Reconsidering conceptual change: Issues in theory and practice*, 3-27.

Chi, M. T., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012b). Misconceived causal explanations for emergent processes. *Cognitive science, 36*(1), 1-61. https://doi.org/10.1111/j.1551-6709.2011.01207.x

Danish, J. A. (2014). Applying an activity theory lens to designing instruction for learning about the structure, behavior, and function of a honeybee system. *Journal of the Learning Sciences, 23*(2), 100-148. https://doi.org/10.1080/10508406.2013.856793

Dickes, A. C., & Sengupta, P. (2013). Learning Natural Selection in 4th Grade with Multi-Agent-Based Computational Models. In *Research in Science Education* (Vol. 43, Issue 3). https://doi.org/10.1007/s11165-012-9293-2

Dickes, A. C., Sengupta, P., Farris, A. V., & Basu, S. (2016). Development of mechanistic reasoning and multilevel explanations of ecology in third grade using agent-based models. *Science Education*, *100*(4), 734-776. https://doi.org/10.1002/sce.21217

Elsawah, S., Ho, A. T. L., & Ryan, M. J. (2021). Teaching Systems Thinking in Higher Education. *INFORMS Transactions on Education.* https://doi.org/10.1287/ited.2021.0248

Fiedler, D., Tröbst, S., Großschedl, J., & Harms, U. (2018). EvoSketch: Simple simulations for learning random and probabilistic processes in evolution, and effects of instructional support on learners' conceptual knowledge. *Evolution: Education and Outreach, 11*(1). https://doi.org/10.1186/s12052-018-0089-3

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive psychology, 15*(1), 1-38. https://doi.org/10.1016/0010-0285(83)90002-6

Goel, A. K., Rugaber, S., & Vattam, S. (2009). Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language. *Ai Edam*, *23*(1), 23-35. https://doi.org/10.1017/S0890060409000080

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *The Journal of the Learning Sciences*, *17*(4), 465-516. https://doi.org/10.1080/10508400802394898

Goods, A. R. (2019). V-Note: a video analysis tool for teacher| researchers. *Cultural Studies of Science Education*, *14*(3), 753-767. https://doi.org/10.1007/s11422-018-9877-x

Gregory, T. R. (2009). Understanding Natural Selection: Essential Concepts and Common Misconceptions. *Evolution: Education and Outreach*, *2*(2), 156–175. https://doi.org/10.1007/s12052-009-0128-1

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive science*, *28*(1), 127-138. https://doi.org/10.1207/s15516709cog2801_7

Holland, J. H. (2000). Emergence: From chaos to order. OUP Oxford.

Horwitz, P., McIntyre, C. A., Lord, T. L., O'Dwyer, L. M., & Staudt, C. (2013). Teaching "Evolution readiness" to fourth graders. *Evolution: Education and Outreach*, *6*(1), 1–12. https://doi.org/10.1186/1936-6434-6-21

Hsiao, J.-C., Chen, S.-K., Chen, W., & Lin, S. S. J. (2022). Developing a plugged-in class observation protocol in high-school blended STEM classes: Student engagement, teacher behaviors and student-teacher interaction patterns. *Computers & Education*, *178*, 104403. https://doi.org/10.1016/j.compedu.2021.104403

Jacobson, M. J. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity*, *6*(3), 41-49. https://doi.org/10.1002/cplx.1027

Kelter, J., Peel, A., Bain, C., Anton, G., Dabholkar, S., Horn M. S., Wilensky, U. (2021). Constructionist Co-design: A Dual Approach to Curriculum and Professional Development. *British Journal of Educational Technology*, *52*(3), 1043-1059. https://doi.org/10.1111/bjet.13084

Klopfer, E., Yoon, S. & Um, T. (2005). Teaching Complex Dynamic Systems to Young Students with StarLogo. *Journal of Computers in Mathematics and Science Teaching*, *24*(2), 157-178. Norfolk, VA: Association for the Advancement of Computing in Education (AACE). Retrieved December 17, 2021 from https://www.learntechlib.org/primary/p/5537/.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, *6*(4), 586-597. https://doi.org/10.3758/BF03212967

Levy, S. T., & Wilensky, U. (2008). Inventing a “mid-level” to make ends meet: Reasoning between the levels of complexity. *Cognition and Instruction, 26*(1), 1-47. https://doi.org/10.1080/07370000701798479

Mitchell, M. (2009). Complexity: A guided tour. Oxford University Press.

National Academies. Keck Future Initiative: Complex Systems: Task Group Summaries; National Academies Press: Washington, DC, USA, 2009.

National Academy of Sciences (US). Working Group on Teaching Evolution. (1998). *Teaching about evolution and the nature of science*. Joseph Henry Press.

National Research Council Report (2012)

National Research Council (1996) National Science Education Standards.

Nehm, R. H., & Reilly, L. (2007). Biology majors’ knowledge and misconceptions of natural selection. *BioScience*, *57*(3), 263-272. https://doi.org/10.1641/B570311

Peel, A., Sadler, T. D., & Friedrichsen, P. (2019). Learning natural selection through computational thinking: Unplugged design of algorithmic explanations. *Journal of Research in Science Teaching*, *56*(7), 983-1007.

Penner, D. E. (2000). Explaining systems: Investigating middle school students’ understanding of emergent phenomena. Journal of Research in Science Teaching: The Official Journal *of the National Association for Research in Science Teaching*, *37*(8), 784-806. https://doi.org/10.1002/1098-2736(200010)37:8<784::AID-TEA3>3.0.CO;2-E

PhET Interactive Simulations. (n.d.). Natural Selection [Simulation]. University of Colorado Boulder. Retrieved March 17, 2020, from https://phet.colorado.edu/en/simulations/natural-selection

Pope, D. S., Rounds, C. M., & Clarke-Midura, J. (2017). Testing the effectiveness of two natural selection simulations in the context of a large-enrollment undergraduate laboratory class. *Evolution: Education and Outreach*, *10*(1), 3. https://doi.org/10.1186/s12052-017-0067-1

Sailer, M., Schultz-Pernice, F., & Fischer, F. (2021). Contextual facilitators for learning activities involving technology in higher education: The C♭-model. *Computers in Human Behavior*, *121*, 106794. https://doi.org/10.1016/j.chb.2021.106794

Shaffer, D. W. (2017). Quantitative ethnography. Madison, WI: Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In *C. Lang, G. Siemens, A. F. Wise, & D. Gasevic*

(Eds.), Handbook of learning analytics (pp. 175–187). Society for Learning Analytics Research.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. Cognitive psychology, 52(2), 170-194. https://doi.org/10.1016/j.cogpsych.2005.10.001

Stroup, W. M., & Wilensky, U. (2014). On the embedded complementarity of agent-based and aggregate reasoning in students' developing understanding of dynamic systems. *Technology, Knowledge and Learning, 19(*1-2), 19-52. https://doi.org/10.1007/s10758-014-9218-4

Su, M., Cho, J., Chi, M.T.H., Boucher, N., & Vanbibber, B. (2021). Designing Simulation Module to Diagnose Misconceptions in Learning Natural Selection. In *de Vries, E., Hod, Y., & Ahn J. (Eds.), Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021*, pp.410-417. Bochum, Germany: International Society of the Learning Sciences. https://doi.dx.org/10.22318/icls2021.410

Su, M., Cho, J., Vanbibber, B., Ha, J., & Chi, M. (2020). Diagnosis of Misconceptions with Coherent Underlying Structure in Learning Diffusion. In *Gresalfi, M. and Horn, I. S. (Eds.), The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences* (ICLS) 2020, Volume 4 (pp. 2321-2324). Nashville, Tennessee: International Society of the Learning Sciences. https://doi.dx.org/10.22318/icls2020.2321

Su, M., Chi, M.T.H., Qian, Y.X., Boucher, N. (In Preparation). *Designing Simulation Modules to Foster Understandings of Emergent Processes and Address Misconceptions in Learning Natural Selection.*

Wagh, A. & Wilensky, U. (2013). Leveling the Playing Field: Making Multi-level Evolutionary Processes Accessible through Participatory Simulations. In *Rummel, N., Kapur, M., Nathan, M., & Puntambekar, S. (Eds.), CSCL 2013 Conference Proceedings* Volume 2 (pp. 181-184). Madison, WI. https://doi.dx.org/10.22318/cscl2013.2.181

Wagh, A., & Wilensky, U. (2018). EvoBuild: A Quickstart Toolkit for Programming Agent-Based Models of Evolutionary Processes. *Journal of Science Education and Technology, 27*(2), 131–146. https://doi.org/10.1007/s10956-017-9713-1

Wickens, T. D., & Keppel, G. (2004). Design and analysis: A researcher's handbook. Upper Saddle River, NJ: Pearson Prentice-Hall.

Wilensky, U. (1997a). NetLogo Traffic Basic model. http://ccl.northwestern.edu/netlogo/models/TrafficBasic. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. (1997b). NetLogo Ants model. http://ccl.northwestern.edu/netlogo/models/Ants. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. (1997c). NetLogo Wolf Sheep Predation model. http://ccl.northwestern.edu/netlogo/models/WolfSheepPredation. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky U. (2018). Restructuration Theory and Agent-based Modeling: Reformulating Knowledge Domains Through Computational Representations. *Designing Constructionist Futures: The Art, Theory, and Practice of Learning Designs*, 287-300.

Wilensky, U., & Centola, D. (2007, November). Simulated evolution: Facilitating students' understanding of the multiple levels of fitness through multi-agent modeling. In *evolution challenges conference*. Phoenix, AZ.

Wilensky, U., & Jacobson, M. J. (2014). Complex systems and the learning sciences. In *The Cambridge handbook of the learning sciences*, Second Edition (pp. 319-338). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.020

Wilensky, U., & Novak, M. (2010). Teaching and learning evolution as an emergent process: The BEAGLE project. *Epistemology and science education: understanding the evolution versus intelligent design controversy*. Routledge, New York.

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Mit Press. https://doi.org/10.1023/A:1009421303064

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories - An embodied modeling approach. *Cognition and Instruction*, *24*(2), 171–209. https://doi.org/10.1207/s1532690xci2402_1

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and technology*, *8*(1), 3-19. https://doi.org/10.1023/A:1009421303064

Xiang, L., & Passmore, C. (2015). A framework for model-based inquiry through agent-based programming. *Journal of Science Education and Technology, 24*, 311-329. https://doi.org/10.1007/s10956-014-9534-4

Yoon, S. A. (2021). Complex Systems Research in K12 Science Education: A Focus on What Works for Whom and under Which Conditions. *Systems*, *9*(2), 29. https://doi.org/10.3390/systems9020029

Yoon, S. A., Goh, S. E., & Park, M. (2018). Teaching and learning about complex systems in K–12 science education: A review of empirical studies 1995–2015.

Review of Educational Research, *88*(2), 285-325. https://doi.org/10.3102/0034654317746090

APPENDIX A

PRE-POSTTEST

Welcome to the Pretest!

The purpose of the pretest is to understand your knowledge about Natural Selection and other emergent phenomena. It consists of 2 sections and will take you 20-30 minutes to complete:

Section A: You will work on three sets of "True or False" questions. For the questions that are False, you will be asked to provide explanations.

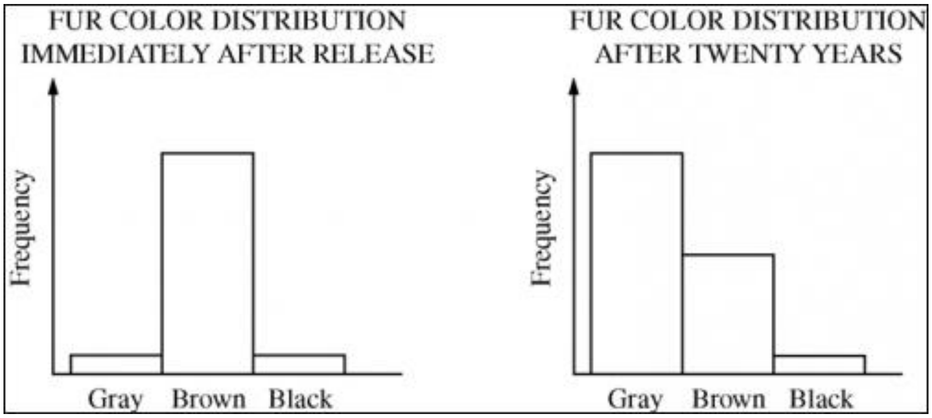
Section B: You will work on four open-ended questions. Two of them are about natural selection, whereas the other two will ask you to explain emergent phenomena that are common in our daily life.

When asked to give an explanation, please try to be elaborative and provide evidence to support your claims. Now, go ahead to continue with the test!

Page Break

Section A:

In this section, decide "True or False" for each statement and provide explanations for the false statements you identified. You should select "False" if you disagree with any part of the statement.



Q1. Decide True or False for each of the following statements that explains the change in the frequency distribution of fur color in the mouse population after 20 years, as shown in the figures above.

Q1a. The number of gray mice in the population increases slightly each year for 20 years, which adds up to the pattern of gray fur becoming more common in the population.

○ True

○ **False (This is the correct answer)**

Q1a_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q1b. Changes you observed in the mice's fur color patterns from the start of the experiment to the end were wholly due to random factors.

○ True

○ **False (This is the correct answer)**

Q1b_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q1c. The proportion of brown mice has decreased in relation to gray mice over the 20 generations, but it is possible that the number of brown mice increased for several generations.

○ **True (This is the correct answer)**

○ False

Q1c_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q1d. The gray mice chose not to reproduce with the brown mice and not to give birth to brown mice because the gray fur color trait was more advantageous.

○ True

○ **False (This is the correct answer)**

Q1d_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q1e. Any mouse can reproduce with any other mice of the opposite sex once they survive to a mature age and are able to reproduce.

○ **True (This is the correct answer)**

○ False

Q1e_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Page Break

Q2. Decide True or False for each of the following statements that describe what happens when a population of bacteria becomes resistant to an antibiotic. Note: a bacterium is one individual in a group of bacteria.

Q2a. During treatment with an antibiotic, each individual bacterium mutates to become resistant. Only the bacteria which become resistant survive.

○ True

○ **False (This is the correct answer)**

Q2a_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q2b. Resistant bacteria only appeared after the population of bacteria is exposed to the antibiotic.

○ True

○ **False (This is the correct answer)**

Q2b_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q2c. The bacteria population gradually become more resistant to the antibiotic every generation the more they are exposed to it.

○ True

○ **False (This is the correct answer)**

Q2c_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q2d. The number of resistant bacteria increases every generation, so resistant bacteria will become more common in the population over many generations.

○ True

○ **False (This is the correct answer)**

Q2d_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q2e. Although some bacteria with the resistant trait do not survive, overall, the resistant trait becomes more common in the population as the ones that survive outnumber the ones that didn't survive.

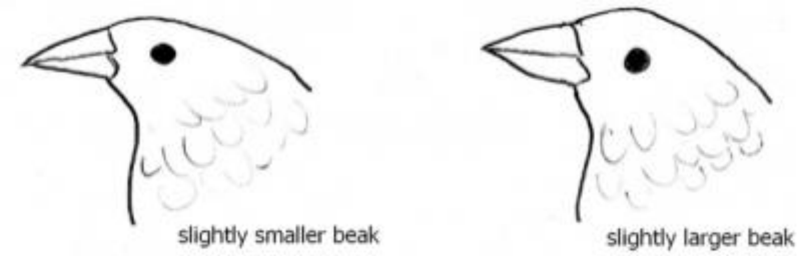
○ **True (This is the correct answer)**

○ False

Q2e_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.
____

Page Break

Q3. A species of bird ate many types of seeds, each type coming from a different species of tree. The birds’ beaks varied in size, with some individuals having slightly smaller beaks and others having slightly larger beaks.



A few years went by without much rain, and the only species of tree that survived had large seeds. Many generations later, almost all the birds had slightly larger beaks.

Decide True or False for each of the following statements that try to explain this phenomenon.

Q3a. The birds with larger beaks were better at eating the large seeds than those with smaller beaks, so the birds with larger beaks will only reproduce with larger beaks and have offspring with beneficial traits.

○ True

○ **False (This is the correct answer)**

Q3a_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.
____

Q3b. At any point in time, birds' beaks becoming larger reflects all the birds’ interactions, such as larger beaks mating with smaller beaks, and larger beaks eat larger seeds more easily than smaller beaks.

○ **True (This is the correct answer)**

○ False

Q3b_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q3c. Most birds with smaller beaks had to work harder than those with larger beaks to crack open the large seeds. The more they used their beaks, the larger their beaks became, so that they would be better able to eat the large seeds and get enough food to survive, reproduce, and pass the trait of larger beaks to the next generation.

○ True

○ **False (This is the correct answer)**

Q3c_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q3d. In all generations, larger beaked birds survived better than shorter beaked birds, so the number of larger beaked birds increased every generation.

○ True

○ **False (This is the correct answer)**

Q3d_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Q3e. In most generations, a greater number of larger beaked birds reproduced than shorter beaked birds, so the larger beaked birds became more common in the population, relative to the short beaked birds, over time.

○ **True (This is the correct answer)**

○ False

Q3e_OE (Display this question if False is selected)
Please explain below why you think the above statement is false.

____

Page Break

Section B:

In this section, please respond to the open-ended questions. Please try to type down at least 5 explanatory sentences for each question.

Q4. How would biologists explain how a living mouse species with claws evolved from an ancestral mouse species that lacked claws?

Q5. How would a biologist explain how a living mosquito species resistant to DDT evolved from an ancestral mosquito species that lacked resistance to DDT? DDT: a powerful insecticide that is poisonous to mosquitos.

Q6. When a flock of geese flies south in a V-formation, and a storm appears, the storm will slow down the flock of flying geese because the wind blows against the birds, but the V is still maintained.

Q6a (Posttest). Please observe the simulation below and explain how the simulation can help you understand the process of geese flying into V-shape.

Q6b (Pretest) Please explain why the V-pattern is still maintained?

Q7. In recent years, fake news has proliferated via social media, in part because they are so easily and quickly shared online. Please observe the simulation below. In the simulation, a node creates fake news and shares it over the network.

Q7a (Posttest). Please observe the simulation and explain how the simulation can help you understand the process of fake news spreading.

____

____

____

Q7b (Pretest). How would you stop fake news from spreading through the internet?

____

____

____

End of Test

APPENDIX B

CROSS-TABULATION OF TEST CONSTRUCT

Cross Tabulation of the Test Construct for Each True of False Item on the Pretest

Construct	Understanding of Emergent Causality in the context of Natural Selection											
	Relations			Inter-level Causal Relationships		Implications		Natural Selection				
Item	F1	F2	F3/ F4	F5	F6	F7a	F7b	DP 1	DP 2	DP 3	DP 4	DP 5
Q1a					X							
Q1b										X	X	
Q1c							X	X				X
Q1d		X				X			X			
Q1e		X	X									
Q2a	X			X								
Q2b			X		X			X				
Q2c					X		X					
Q2d				X								X
Q2e	X									X		
Q3a	X			X		X						
Q3b		X	X									X
Q3c					X			X	X			
Q3d				X			X			X		
Q3e									X	X	X	

PAIR-C Features	Darwinian Principles (DP)
F1: Same Set/ Distinct Set F2: Random/ Restricted F3: Simultanous/ Serial order F4: Independent/ Dependent F5: Converging change/ Incremental change F6: Collective summing/ Cumulative summing F7: No causal agent/ causal agent F8: Not align/ align	DP1: Random intraspecies variability (individual variation) DP2: Heritability of certain traits (or genetic determination) DP3: Differential survival rate (local adaptation) DP4: Differential reproductive rate (reproductive advantage) DP5: Accumulation of changes over many generations

This table provides a cross-tabulation of the construct for each item. Since responding to each item typically involves an understanding of PAIR-C features and DP, most of the items used in the test appear in multiple categories, and no two items assess the same combination of categories.

186

APPENDIX C

OVERVIEW OF PRE-POSTTEST ITEM

Test Item Overview

No.		**Difficulty**	**Source**	**History**	**Format**
Q1	Q1a	Deep	College Board's AP Biology Exam 2013	Used in Pilot Study 1 and 2	Two-tiered True or False questions
	Q1b	Shallow			
	Q1c	Shallow			
	Q1d	Shallow			
	Q1e	Shallow			
Q2	Q2a	Deep	AAAS	Used in Pilot Study 1 and 2	
	Q2b	Deep			
	Q2c	Deep			
	Q2d	Deep			
	Q2e	Shallow			
Q3	Q3a	Deep	AAAS	Used in Pilot Study 1 and 2	
	Q3b	Shallow			
	Q3c	Deep			
	Q3d	Deep			
	Q3e	Shallow			
Q4		Deep	Peel et al. (2019)	n/a	Open-ended questions
Q5		Deep		n/a	
Q6		Deep	Learning and Cognition Lab	Used in Pilot Study 1 and 2	
Q7		Deep	Learning and Cognition Lab	Used in Pilot Study 2	

APPENDIX D

LEARNING MODULE PROMPT QUESTIONS

Simulation Module 1: Peppered Moths Scenario		
Activity	Prompt Questions	ICAP
Activity 1	1. Describe what you see in this simulation environment.	A
	2. Before running the simulation, please predict how the percentage of moths with each color will change over time under the current circumstance?	C
Activity 2	3. Observe the simulation and describe what you observed in the simulation world from tick 1 to tick 100.	A
	4. What information on this page did you miss in your response?	C
	5. Select the diagram below that best describes what will happen in terms of the distribution of color variations among moths over 100 ticks? What information on this page is most important for you to understand?	A-C
	6. Watch a simulation video. Describe what you notice when the selection pressure is at 0 versus the selection pressure is at 100?	A
	7. Run the simulation and explain how the distribution change of color variations is influenced by selection pressures.	C
	8. What information on this page is most important for you to understand?	C
Activity 3	9. Reset the simulation ("selection" slider to 30, pollution level to be around 50%). Run the simulation till around tick 100. Describe what is happening in the simulation world.	A
	10. What information on this page did you miss in your response?	C
	11. Do NOT reset the simulation. Continue clicking the "pollute" button to set the pollution level at 80%. Now, run the simulation till tick 200 and describe what you observed.	A
	12. What do you learn from this simulation video?	C

	13. Explain how the color variations at Tick 125 is different from Tick 0.	C
	14. Explain why Tick 0 looks different from Tick 125?	C
	15. What information on this page is most important for you to understand?	C
Activity 4	16. Decide True or False for the following statement: If a newborn moth has a mutation, its color will always be darker than its parents. Please explain your response in 2-3 sentences.	A-C
	17. What information on this page is most important for you to understand?	C
	18. Explain below how the pattern changes over time?	C
	19. What information on this page is most important for you to understand?	C
	20. Observe the data table provided by a researcher. Please explain why it occurs that the number of dark-colored moths drops at Generation 5 from 229 to 220 while the percentage of dark moths increases over time from 61% to 67%?	C
Simulation Module 2: Rock Pocket Mice		
Activity	Prompt Questions	ICAP
Activity 1	21. Press the “Setup” button of the simulation. Describe what you see in this simulation environment.	A
	22. Change the sliders under the "Initial Settings" in the simulation. Try to change the settings such that all the mice have light-colored fur. Once you get all mice with light fur, describe the initial settings you used.	A
	23. Predict what will happen after lots of generations if the initial population of mice all has light-colored fur?	C

	24. Run an experiment to prove or disprove your answer to the previous question and explain your observations.	C
	25. There are two alleles "A" (dominant) and "a" (recessive). Based on your investigations so far, can you say what would the phenotype (fur color: light or dark) be of the three genotypes "AA", "Aa", and "aa"? - AA Please explain how you figured out the answer to the previous question.	A-C
Activity 2	26. What did you notice from the simulation video?	A
	27. Please explain how the variation in the mice population originated in the real world?	C
	28. What information on this page is most important for you to understand?	C
	29. Start the simulation with a mixed population: some mice with dark fur and some mice with light fur. Run the simulation and describe how the distribution change of color variations is influenced by the initial settings.	A
	30. What else did you notice from the simulation video?	C
	31. What information on this page did you miss in your response?	C
Activity 3	32. Run the simulation and describe what do you notice about how the mice population changes over time with the dark background?	A
	33. Please explain why the changes happen?	C
	34. What information on this page is most important for you to understand?	C
	35. Please describe what do you notice about how the mice population changes over time with the different backgrounds?	A

	36. If there is a mixed population of mice evenly distributed in a dark background environment, as time goes on, mice with dark fur become more common in the population. Please explain how the changes happen?	C
	37. What information on this page did you miss in your response?	C
	38. a student tried to come up with some general statements to elaborate on the process of natural selection. Please help this student to re-organize these statements in the correct order by dragging and dropping each statement into the correct order.	A
	39. What information on this page is most important for you to understand?	C
Activity 4	40. Please select three misconception statements you'd like to provide explanations with. This statement is misconceived because: This is a misconception because: This statement is misconceived because:	C

APPENDIX E

QUESTIONNAIRES

Background Questionnaire:

Q1. What is your full name (First name, Last name)?

____

Q2. To which gender identity do you most identify?

○ Male

○ Female

○ Non-binary

○ I prefer not to answer

Q3. What is your age?

____

Q4. Which of the following best represents your racial or ethnic heritage? Choose all that apply.

☐ White

☐ Black or African American

☐ Hispanic/Latino

☐ American Indian/Alaska Native

☐ Asian

☐ Hawaiian Native or Other Pacific Islander

☐ Other

☐ Do Not Wish to Say

Q5a. Is English your first language?

○ Yes

○ No

Display This Question:

If 5a. Is English your first language? = No

Q5b. If English is not your first language, what is your first language?

____

Q6. Please indicate your educational background.

○ Undergraduate, Freshman

○ Undergraduate, Sophomore

○ Undergraduate, Junior

○ Undergraduate, Senior

○ Graduate students, Master level

○ Graduate students, Doctoral level

○ Other ____

Q7. What is your major/program at ASU?

____

Q8. How many biology classes have you taken in high school?

____

Q9. Have you heard about and used NetLogo simulation before?

○ Yes, I have heard about and used NetLogo simulation before the study.

○ I have heard about it but never used NetLogo simulation.

○ No, I have never heard about nor used NetLogo simulation.

○ I don't remember.

Q10. Are you willing to participate in a 20-minute interview after the study?

○ Yes

○ No

○ Not sure yet

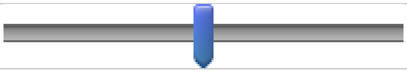
Post-Project Questionnaire:

Q1. How engaging do you feel about this simulation unit?

Not Engaging Very Engaging

0 1 2 3 4 5

.

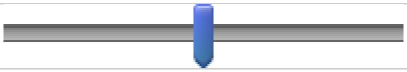


Q2. Is this simulation helpful for understanding natural selection?

Not Helpful Very Helpful

0 1 2 3 4 5

.

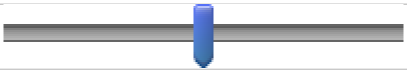


Q3. Are the instructions and prompt questions clear to you in this simulation unit?

Not Clear Very Clear

0 1 2 3 4 5

.



Q4. Was there anything hard to understand or confusing to you in the simulation unit? If Yes, could you give us an example?

____

____

Q5. How is the NetLogo simulation similar or different from other simulations you've played before?

____

____

Q6. Approximately, how much time did you spend on each simulation scenario?

○ Peppered Moths ____

○ Rock Pocket Mice ____

Q7. In general, what do you like or dislike about the Technology in Action Project?

____

____

Q8. Are you willing to participate in a 20-minute interview after the study?

○ Yes

○ No

If Q8 = Yes, then display the following:

Thank you so much for willing to participate in the 20-minute interview!
During the 1:1 interview, the researcher will ask you some simple questions to learn more about your interactions with the simulations and your overall learning experience. You can also use this opportunity to ask any questions related to the project.
Please click this link https://calendly.com/mansu2022/20min to choose your preferred interview time. You will receive an email notification and a Zoom invitation link once the interview time is confirmed.

Congratulations! You've reached the end of the Technology in Action project!
Please click the "Next" button to submit your response!

APPENDIX F

POST-STUDY INTERVIEW PROTOCOL

Interview Protocol
Interview Type: Semi-structured Clinical Interview
Time: 20-30 minutes, one session

Hello, I'm Man Su. I'm a researcher at ASU Learning and Cognition Lab. Thank you so much for signing up for the interview and willing to share more about your experience with the Technology in Action project. In this interview, I am going to ask you some questions to better understand your responses to the surveys and to have you reflect more on your overall learning experience. By the end of the interview, you can also ask me any questions about the project.

First, is it okay with you if I start recording now?

Before the experiment, you filled out a survey in which you mentioned that you major in [MAJOR]

- **Q1: As a [MAJOR] student, do you have any experience using simulations in your own learning before this project?**
    - If Yes, How often do you use simulations (or other multimedia resources) in your classes (online and in-person)?
    - *[Comparing Different Simulations] In your post-survey, you mentioned that [..........] So, what are the other simulations you've used before? In what circumstances did you use other simulations?*
    - In general, how do you feel about learning through simulations in an online course? Versus in-person courses?
- **Q2: Could you tell me something more about your major?**
    - If Major in Education
        - What subjects do you teach?
        - As a pre-service [SUBJECT] teacher, do you use or plan to use simulations in your own (future) teaching?
        - Could you describe how you will use simulations in your own teaching?

- **Q3: In the survey, you've also mentioned that you had taken [# biology classes in high school].**
    - What does a typical biology lesson look like to you?
        - Do you use simulations? How often do you use simulations?
        - Have you ever used a simulation when learning about [Natural Selection] before the experiment?

During the experiment, you've experienced two simulation scenarios that explained Natural Selection

- **Q4: In what ways, if any, has your understanding of Natural Selection changed since the start of the project?**
    - How would you explain natural selection differently to others?

    - At the beginning of each simulation scenario, you've raised some questions about natural selection phenomena. Were your questions answered by the end of the simulation units?

- **Q5: I will ask you some questions about your use of simulations in this project.**
    - In what ways, if any, has your interactions with simulations changed since the start of the project?
        - What do you think about the difficulty of the second simulation scenario when compared to the first scenario?
    - Do you find the videos and screenshots helpful visual representations to explain natural selection? Why or why not?
        - For the PAIR-C group: Did you notice the blue and red links in the simulation videos? What do you think about using these links to visualize agent interactions? [show the Simulation]
        - Did you adjust the playback speed of the embedded video?
        - Did you replay any of the simulation videos?
    - What support (e.g., instructional or technical support) do you wish you had during the simulation scenarios?

After the experiment, you watched two other simulation videos in the posttest. One is about Geese flying in a V shape and the other one is about fake news spreading.

- **Q6: How confident do you feel about using simulations to understand or explain these complex concepts?**
    - Do you think what you learned in the natural selection simulation scenarios can be applied to understanding other new concepts?
    - Did you notice any commonalities between these phenomena and the natural selection process?

Do you have anything else you want to say or ask about this project?

Great, well thank you so much for participating. I'm going to stop recording now.

APPENDIX G

SCORING RUBRICS

Q1a. The number of gray mice in the population increases slightly each year for 20 years, which adds up to the pattern of gray fur becoming more common in the population.	
Q1a_OE. *Feature 6 collective summing* This statement is **False**. ● "The population of grey mice ***could not possibly increase every single year***, this falls under "***does not align***". ● "One year there could be a surge of gray mice and one year there could be many dying instead. It is ***not a continual increase***." ● "Numbers ***do not necessarily increase in each generation.***"	If the explanation mentions or indicates that the increase is not gradual or continuous by every generation, then give 1 point;
● Both graphs ***do NOT provide enough info*** to make the statement above. The graph on the right looks to be showcasing a decrease over time. ● While the grey mice population did increase significantly, ***without more data we don't know*** that it increased slightly for 20 years.	if only refers to not enough data or info, give 0.5 point;
● "because the grey mice population grew the most dramatically" ● "The population of mice with grey fur dramatically increases." ● I don't know if the population increases slightly, or if that is why they end up more common.	if the answer only describes the bar chart, give 0 point.
Q1b. Changes you observed in the mice's fur color patterns from the start of the experiment to the end were wholly due to random factors.	
Q1b_OE. This statement is **False. (DP3 DP4)** ● "This is tricky. It is somewhat random I do agree, but some mice do ***have a better chance of surviving compared to others***." ● "Changes observed were ***affected by mutations and by mice-bird interactions.***" ● "Although mutation is technically a random factor, ***operations of genetics, adaption, predation, mating, etc. influenced the observable changes*** of the experiment."	If indicates interactions (mating interactions, predator-prey interactions) or considers survival and reproductive advantages, give 1 point;
● "The changes were due to the ***traits the genes express***." ● "They were more likely ***due to environmental factors or gene-dominance***." ● "Changes in the mice's fur color patterns were ***due to mutation***."	if only state environmental factors or mutation factors without considiering interactions, give 0.5 point;
● not enough information on the chart. ● did not specify other factors. ● "The factors weren't random. We changed the environment, and that is what determined what happened." ● "The fur changes because of the environment around them."	if the answer contains misconception, give 0 point.

● “Depending on number of predators and background color advantages, certain mice were able to ensure their survival.”	
Q1c. The proportion of brown mice has decreased in relation to gray mice over the 20 generations, but it is possible that the number of brown mice increased for several generations. Q1c_OE. This statement is **True.** [Feature 7b: not align; DP1: individual variation; DP5: accumulated change over time]	
Q1d. The gray mice chose not to reproduce with the brown mice and not to give birth to brown mice because the gray fur color trait was more advantageous.	
Q1d_OE. This statement is **False**. [Feature 2: random; Feature 7a: no causal agent; DP2: heritability] because it assumes that gray mice have special status and intentionally restrict their mating interactions within a subgroup that has the same color trait and exclude the possibility that a gray mouse can either reproduce with another gray mouse or a brown/black one. It also misinterprets differential reproductive advantage as applied to subgroups rather than the entire collection of mice population and wrongly assumes that a subgroup of fit mice only reproduces with other fit ones. The mating interaction is a **random** process. ● “This is not true because the ***mice do not choose who they mate with it is random***.” ● “Mice ***don't pick and choose traits***” ● “Mice ***do not pick who to reproduce with based off of fur color***.”	If mention any of the highlighted reasons (in red ink), give 1 point;
● “You don't know for certain that they chose not to because of the color of their fur. There are ***other factors*** to consider.” ● “There is ***not enough data*** to prove that from the two graphs alone.”	Only vaguely explained or referred to not enough data from graphs, give 0.5 point;
● “I believe they mated with the brown mice in order to increase their numbers and some unrevealed factors influenced the advantage of gray fur.”	if the answer contains misconception or irrelevant, give 0 point.
Q1e. Any mouse can reproduce with any other mice of the opposite sex once they survive to a mature age and are able to reproduce. Q1e_OE. This statement is **True.** [Feature 2: random Feature 3&4: simultaneous and independent]	
Q2a. During treatment with an antibiotic, each individual bacterium mutates to become resistant, and all these individuals survive to pass this trait to their offspring. Only the bacteria which become resistant survive.	

Q2a_OE. This statement is **False**. (Feature 1: distinct set; Feature 5: incremental change) because it assumes that the subgroup of resistant bacteria all survive whereas unresistant bacteria all die. This misinterprets differential survival advantage as applied to subgroups and wrongly believes that resistant bacteria incrementally produce the pattern of resistant bacteria becoming more common by passing the advantageous traits to their offspring. ● “***Bacterium can survive that are not exposed to an antibiotic*** or exposed to a weaker antibiotic. This also depends on its inherited traits.” ● “This is false because the ***bacterium cannot become resistant in one generation*** it takes time.” ● “There is always a chance that ***bacteria which are not resistant could still survive***.”	If mentions or indicates any of these reasons, give 1 point;
● “the offspring will become resistant too.” ● “I don't believe every bacterium mutates. Just the resistant ones.”	The answer is correct but vague, give 0.5 point;
● “the resistant bacteria survive and the ones that are not resistant die. over time the resistant bacteria reproduce to create more resistant bacteria.” ● “Each individual bacterium does not mutate - just one needs to and pass on their mutations.”	if the answer contains misconception or irrelecant, give 0 point.
Q2b. Resistant bacteria only appeared after the population of bacteria is exposed to the antibiotic.	
Q2b_OE. This statement is **False. (Feature 3: Serial, DP1, collective summing)** because it misunderstands that bacteria can only become resistant after they are exposed to antibiotics. In fact, bacteria can already have resistant traits before they are exposed to antibiotics, rather than “inventing” new traits in order to fit the environment. Resistant bacteria appeared due to the net effect of bacteria surviving and dying when exposed to the antibiotic. ● “There could be a mutation that allows the ***bacteria to be resistant to the antibiotic with a prior exposure***.” ● “They have had mutations in the bacteria that allowed them to b***e more resistant prior to exposure***.”	If indicates the resistant traits exist before exposing to antibiotics give 1 point;
● “It could ***mutate,*** but adding the antibiotic introduces it faster.” ● “The bacteria ***just mutate to become stronger and more resilient***. I do not believe they became more resistant after becoming exposed to the antibiotic.”	if only state mutation factors without providing explanations, give 0.5 point;
● “Antibiotic helps with the cure of the bacteria.”	if the answer is irrelevant, give 0 point.

Q2c. The bacteria population gradually become more resistant to the antibiotic every generation the more they are exposed to it.	
Q2c_OE. This statement is False. (Feature 6: cumulative summing, Feature7b: align) because it assumes that the traits will become more resistant as they are more exposed to antibiotics. However, some bacteria can be less resistant as the agent-level properties do not always align with the overall pattern. ● "***Not every generation*** will have the same results of becoming resistant."	If mentioned about inconsistency or not align, give 1 point;
● "After mutating, bacteria will ***pass their favorable traits to the next generation*** and those traits ***might make them more resistant***." ● "Some generations ***could remain the same amount of resistant***."	If answers did not explain the causal mechanism, and only stated facts or observations, give 0.5 point;
Q2d. The number of resistant bacteria increases every generation, so resistant bacteria will become more common in the population over many generations.	
Q2d_OE. This statement is False. (Feature 5: converging pattern, DP5) because it assumes the number of resistant bacteria at any point in time is increasing and is added to the previous generation; it implies that over time the pattern of resistant bacteria arises and is a predictable pattern, however one cannot predict what the interim pattern would look like. ● "Each generation ***fluctuates in numbers*** of resistant bacteria." ● "I think the resistance of bacteria increases in general and ***not the number*** of bacterium that become resistant." ● "***Not guaranteed to increase*** every generation."	If answers mentioned or indicated numbers not always increasing, give 1 point;
● "Truthfully I am not sure of this so I cannot say true. I think ***there are other factors and different mutations that could change this*** overtime."	If answers were correct but vague, give 0.5 point;
● "I think the over use of antibiotic in one body might the cause to drug-resistant bacteria. I don't think or know if bacteria passes to another generation."	if the answer was irrelevant, give 0 point.
Q2e. Although some bacteria with the resistant trait do not survive, overall, the resistant trait becomes more common in the population as the ones that survive outnumber the ones that didn't survive. Q2e_OE.	

This statement is **True.** [Feature 1: same set, DP3: survival advantage]	
Q3a. The birds with larger beaks were better at eating the large seeds than those with smaller beaks, so the birds with larger beaks will only reproduce with larger beaks and have offspring with beneficial traits.	
Q3a_OE. This statement is **False**. (Feature 1: distinct set, F7a: control agent) ● "There is still the factor of the ***recessive trait***. Therefore, the birds with larger beaks could still have smaller beaked offspring based on their mate." ● "This is not true because birds ***do not pick*** on who they mate with. It is at ***random***." ● "There is still the possibility that the offspring would have a smaller beak depending of ***who the bird mated with and what their genotype*** may be."	If mention any of these reasons (random, genotype, traits), give 1 point;
● "Larger beaked birds ***do not just mate with*** larger beaked birds." ● "Larger beaked birds ***could/would still reproduce with*** smaller beaked birds."	if only state larger beaked birds still mate with other birds without providing explanations, give 0.5 point;
● "No, since the only species of tree that was left was large seed tree. The bird ***evolved to protect their species to only have large beaks*** for more efficiency and survival." ● "The seed size's had nothing to do with the size of their beaks."	if the answer contains a misconception or irrelevant, give 0 point.
Q3b. At any point in time, birds' beaks becoming larger reflects all the birds' interactions, such as larger beaks mating with smaller beaks, and larger beaks eat larger seeds more easily than smaller beaks. This statement is True. (Feature 2: random, Feature 4: independent, DP5)	
Q3c. Most birds with smaller beaks had to work harder than those with larger beaks to crack open the large seeds. The more they used their beaks, the larger their beaks became, so that they would be better able to eat the large seeds and get enough food to survive, reproduce, and pass the trait of larger beaks to the next generation.	
Q3c_OE. This statement is **False**. (Feature 6: cumulative summing, DP1&2)	If mentioned or indicated unintentional, or due to traits, give 1 point;

because it assumes that birds can intentionally transform and develop new traits in order to better fit the environment within one generation and pass the advantageous trait to their offspring. ● “Their beaks ***didn't grow out of use***. They would have grown ***due to inheriting traits***.” ● “There is ***no way for a birds beak to grow during its lifetime***. It would need to be ***due to a genetic mutation*** and when the birds with the mutation were able to get more food and survive, they passed on the gene.” ● “The ***environment does not effect the reproduction, but the dominant and recessive traits do***.”	
● “The bird's beak size **didn't change**” ● “Birds with smaller beaks **could not just get bigger beaks** from eating nuts.” ● “An organisms physical traits **cannot naturally change**.”	if only state it didn’t change without providing explanations, give 0.5 point;
● “I believe the birds with smaller beaks were unable to survive the situation, while the birds with larger beaks could eat and move on to breed.” ● “The bird's beak size didn't change - the birds with the bigger beaks survived and those with smaller beaks died.”	if the answer contains misconception, give 0 point.
Q3d. In all generations, larger beaked birds survived better than shorter beaked birds, so the number of larger beaked birds increased every generation.	
Q3d_OE. This statement is False. (Feature 5 incremental change, Feature 7b, DP3) because it assumes that the subgroup of larger beaked birds survives better than smaller beaked birds at all times. This misinterprets differential survival advantage as applied to subgroups and wrongly believes that the number of larger beaked birds always increases to be align with the pattern of larger beaked birds becoming more common. ● “Even though the larger beaked birds survived better, ***they wouldn't necessarily increase in population for every generation***.” ● “It is important to remember that number of birds with the desired trait ***does not necessarily increase in every generation***, sometimes it will ***fluctuate***.”	If mentioned or indicated not always increase or “not align”, give 1 point;
● Not in ALL generations, and every year. Most yes, I just do not agree with ALL and EVERY.	Student only focused on the semantic error, give 0.5 point;
● “In all generations after the tree population changed.”	if the answer is irrelevant, give 0 point.

Q3e. In most generations, a greater number of larger beaked birds reproduced than shorter beaked birds, so the larger beaked birds became more common in the population, relative to the short beaked birds, over time.

Q3e_OE.

This statement is True.

[DP3: survival advantage DP4: reproductive advantage]

Q4. How would biologists explain how a living mouse species with claws evolved from an ancestral mouse species that lacked claws?

Key:
The biologists would explain that

- Trait variation (e.g. mouse with claws or lacked claws) in the mice population is present in ancestral time or prior to the introduction of specific selection pressure.
- Mice that survive to adulthood are likely to reproduce and pass on those traits to their offspring.
- From reproduction, random mutations might occur, resulting in variation among offspring.
- Mice with favorable or unfavorable traits who survived would reproduce and pass on those traits to their offspring.
- When selective pressure is introduced, mice with favorable traits (have claws) for the environment would have more chances for survival compared to those with unfavorable traits (lack claws).
- Over time, the population is mainly composed of living mice with claws.

General Rubric:

- A score of "5" should be given to responses that mention
    - "trait variation present prior to selection pressure";
    - "random mutations or trait variations";
    - "survive and reproduce";
    - "influence of selective pressure or envioronment on survival and reproduction rate";
    - "changes occur over time or multiple generations".
- A score of "1-4" should be given to responses that contain 1 to 4 key points that mentioned above.
- A score of "0" should be given to responses that contain no key point or only have misconceptions.

Sample Response	Score	Notes
The mouse ancestor had a mutation that caused him to grow claws./ When it was time, he reproduced./ During reproduction he passed on his traits to his offspring./ The favorable trait was present in the mouse./ Over time, more mice had this trait that was helpful for environmental conditions/ and had more chances to survive./	5	"trait variation present prior to selection pressure"; "mutations occurs"; "survive and reproduce"; "influence of selective pressure or envioronment on survival and reproduction rate"; "changes occur over time".
At one point in time, a mutation within the genome of unclawed mice occurred./ The resulting offspring were able to escape prey faster and dig deeper, safer tunnels. /Over time, these mice produced more young than those without claws and more of their young	4	"mutation occurs" "survive and reproduce offspring" "favorable traits leads to more survival rate" "occurs over time"

made it to maturation and successfully outcompeted the next generation of unclawed mice./ Eventually, only clawed mice were able to survive./		
Claws made life for mice easier, so mice with mutations in their claws /were better able to survive/. Since they were better able to survive, they passed on their genes/and claws became more prevalent./	3	“random mutations or trait variations”; “survive and reproduce”; “higher survival and reproduction rate”.
The mouse with claws had conditions that allowed for the dominant trait to become more common/ and pass the trait on to a point where a mouse that lacked claws was recessive/. Over time the recessive trait was no longer passed on./	2	“survive and reproduce”; “trait variations”. Misconception: “teleological”
They may have mated with another species of mice that had claws./ They could also have adapted for survival./	1	“survive and reproduce”
The need for claws became essential to the mice./ so the ancestors started to grow claws to support their environment./ The mice that started to have claws survived./	0	Misconception: “teleological” Misconception: “essentialism”
It starts off with ancestral mice having to work harder to use their claws./ This then evolved them to get better claws by having to work harder./ Their gene can then can be passed down to their babies./ Which will lead to the mouse species having evolved claws./ They have to live long enough to pass down their gene./	0	Misconception: “use and disuse” Misconception: “teleological”

Q5. How would a biologist explain how a living mosquito species resistant to DDT evolved from an ancestral mosquito species that lacked resistance to DDT? DDT: a powerful insecticide that is poisonous to mosquitos.

Key:
The biologists would explain that

- Trait variation (e.g. mosquito with resistance to DDT or lacked resistance to DDT) in the mosquito population is present in ancestral time or prior to the introduction of specific selection pressure.
- Mosquitos that survive to adulthood are likely to reproduce and pass on those traits to their offspring.
- From reproduction, random mutations might occur, resulting in variation among offspring.
- Mosquitos with favorable or unfavorable traits who survived would reproduce and pass on those traits to their offspring.

- When selective pressure is introduced, mosquitos with favorable traits (have resistance to DDT) for the environment would have more chances for survival compared to those with unfavorable traits (lack resistance to DDT).
- Over time, the population is mainly composed of living mosquitos with resistance to DDT.

General Rubric:

- A score of “5” should be given to responses that mention
    - “trait variation present prior to selection pressure”;
    - “random mutations or trait variations”;
    - “survive and reproduce”;
    - “influence of selective pressure or envioronment on survival and reproduction rate”;
    - “changes occur over time or multiple generations”.
- A score of “1-4” should be given to responses that contain 1 to 4 key points that mentioned above.
- A score of “0” should be given to responses that contain no key point or only have misconceptions.

Sample Response	Score	Notes
The mosquito ancestor had a mutation that affected its genotype./ When it was an adult, it reproduced./ He passed his favorable traits to his baby mosquito./ After selective pressures appeared, the mosquito with the favorable traits survived./ Over time, more mosquitoes had favorable traits./ They were resistant to DDT./	5	“random mutations affected genotype”; “survive and reproduce”; “trait variation present prior to selection pressure”; “influence of selective pressure on favorable traits”; “changes occur over time”.
The mosquito species likely built up a resistance over time./ The mosquitos that survived would have passed on their traits to their offspring./ As each generation reproduces, more offspring are born with the trait of resistance to DDT./ Although, not all of the mosquitos would be born resistance, the population can increase over time./	4	“changes occur over time”; “survive and pass traits”; “trait variation”; “influence of selective pressure on reproduction rate”.
As the mosquito evolved, the future generations developed mutations that made the mosquitos resistant to DDT./ Since those resistant mosquitos were more likely to survive to sexual maturity, the resistant mosquitos were more likely to pass on their advantageous traits to their offspring./ Mosquitos are likely to reproduce with whatever other mosquitos happen to be around them, so they are more likely to come into contact and mate with mosquitos that survive the pesticide./	3	“mutations or trait variations”; “survive and reproduce”; “influence of selective pressure or envioronment on survival and reproduction rate”.

During time, a mosquito that was introduced to ddt may have been resistant./ When that mosquito mates with another who was not resistant, it created a gene that was passed down./ Through generations, more mosquitos will have this new gene of resistance to ddt./ Further along the line mosquitos will be born with the resistance to ddt./	2	“trait variation present prior to selection pressure”; “survive and pass traits”; Misconception: “align”
The resistant could be a recessive trait./ Or the DDT could not be made for the recessive trait./ The DDT could not have known about the ancestral mosquito species./ Over use of the DDT./	1	“trait variations” Misconception: “anthropomorphic”
The longer a mosquito species is exposed, the higher the likelihood that the species gains resistance to DDT due to exposure./	0	Misconception: “use and disuse”

Q6a.

In recent years, fake news has proliferated via social media, in part because they are so easily and quickly shared online. Please observe the simulation below. In the simulation, a node creates fake news and shares it over the network.

Please observe the simulation and explain how the simulation can help you understand the process of fake news spreading.

Key:

- Fake news spreading is an emergent process. The simulation can help students see some emergent features of fake news srpeading.
    - The process is random
    - Spreading fake news can occur simultaneously and without depending on prior outcomes
    - No one could control fake news from spreading
    - The process is not intentional
    - One cannot predict the speed or the scope of the process.

General Rubric:

- Students could earn 1 point if their responses indicate any of the above features or other emergent properties.
- Students could earn a maximum of 5 points for this question.

Sample Response	Score	Notes
It is almost like a disease, how it spreads from site to site./ As soon as one website gets a hold of a story, it becomes impossible to stop the spread of it. /As soon as one gets removed. 5 can appear seemingly from nowhere./	3	The students compared the emergent process of fake news spreading with disease spreading indicates they find some common properties of these two emergent processes. + 1 “impossible to stop” indicates “no control” + 1 “one gets removed, others can appear” indicates “ no dependence on prior outcomes” + 1
This simulation helped me understand how fake news is	2	Compared fake news with wildfire + 1 “can’t stop it…” indicates “no control” + 1

spreading because it literally starts with one and then spreads like a wildfire./ You can't stop it from happening because once one person gets it then another will./		
Fake news spreads through different sources hearing the same story/ and copying the story/ and spreading the fake news./	1	"hear same story, copy the story, spread the fake news" indicate "same set of interactions" + 1
This simulation helps you understand how fake news is spreading because we see once source in the middle and it spreads to several directions.	0	Only describe the pattern shown in the simulation
Fake news all originates from the same source, it is NOT random.	0	Misconception: "NOT random" In the emergent process of fake news spreading, there will be variations of the fake news that originate from the initial source. Therefore, over time the fake news might even be different from the one at the beginning.

Q6b. How would you stop fake news from spreading through the internet?

Key:

- It is difficult to stop the fake news from spreading because there is no one in charge! (Application of the first indicator-Absence of a controlling agents).
- One solution would be to cut the source.
- Another solution would be to identify individuals who are influential or commonly spreading the fake news. By asking these "super-spreaders" to stop spreading the fake news, there is a better chance that the number of people talking about the fake news will go down.
- Another solution would be to foster media literacy among internet users, especially emphasizing the importance of fact checking among youth and adolescents who are easily susceptible to fake news. When more agents change from susceptible or believers to fact-checkers, fake news will become less likely to go viral on the internet.

General Rubric:

- Students could earn 1 point if their responses indicate any of the above solutions or other solutions.
- Students could earn a maximum of 3 points for this question.

Sample Response	Score	Notes
You could stop fake news from spreading by chopping it off at the source./ The spread of misinformation spreads by one person passing it on to another and so forth. You can stop this spread by stopping it in its tracks./ By not allowing this information to spread, you can stop this spread. Doing your own research can stop this and encouraging others to do the same as well./	3	● "chop it off at the source" + 1 ● "Stop the spread by stopping one person passing it to another…" indicates stopping the spreading process by considering or intervening the spreading interactions. + 1 ● "Doing your own research…" indicates "fact checking" + 1

Stronger filters and fact checking	2	● "stronger filters" indicates intervening the spreading interactions or using filters to identify and stop potential super-spreaders. +1 ● "fact checking" + 1
Fact check. Read the material before I share. Read the article. View the source and make sure it is reliable.	1	● "fact checking" + 1
Find the source	0	● Did not explain what to do after finding the source.

Q7a. Please observe the simulation below and explain how the simulation can help you understand the process of geese flying into V-shape.

Key:

- Geese flying into V pattern is an emergent process. The simulation can help students see some emergent features from geese V-formation.
    - Geese can join, leave, or rejoin a V-pattern anytime.
    - No boss agent(goose) control others' positions
    - The process is not intentional, geese cannot have a unified goal of flying towards the same direction
    - The initial random pattern looks different from the final converging pattern of V
    - Not alignment, geese can fly into many different directions even though they are forming a V

General Rubric:

- Students could earn 1 point if their responses indicate any of the above features or other emergent properties.
- Students could earn a maximum of 4 points for this question.

Sample Response	Score	Notes
Flying in a "V" allows for geese to make their groups more aerodynamic./ The geese can be "seen" trading places at the end of the V to take turns riding in the easiest positions within the flying pattern./ The flying at the beginning of the video is erratic and chaotic, the goose would have to change its flight pattern constantly reacting to those around it./ With flying in a "V", geese can fly with less change in direction and conserve energy./	3	● "trading places" indicates "no boss goose controlling places or positions" +1 ● "at the beginning is erratic and chaotic, flight pattern react to those around it…" indicates the converging mechanism of pattern formation. + 1 ● "less change in direction and conserve energy" indicates local goals rather than global goals of purposefully staying at one direction etc. + 1
This simulation can help me understand why geese fly in a V-shape because it shows them all over the place/ and with time, they follow one another to stay safe and not run into each other./	2	● "all over the place" indicates random emergent process + 1 ● "they follow one another and not run into each other" indicates same sets of interactions + 1
When flying in a V shape they are moving at a much faster rate than before. /They are also being able to	1	● "fly at faster rate, move without crashing" indicates same sets of interactions each goose would have + 1

move without crashing into each other in every different direction./		
It seems like the geese finds a strong leader to follow./ However, there could also be strong winds. I think the geese fly in different directions to see where the wind is the strongest and how they can fly against/with it.	1	● Misconception: “strong leader” – “controlling agent” ● “fly in different directions to adjust to the wind” indicates local goals rather than global goals + 1
Once the birds realize they had the same goal they then formed a uniformed oder in which they fly in.	0	● Misconception: birds realize that they had the same goal – “global goal” ● Misconception: Then they formed a uniformed order – “sequentail order”, “teleological”, or “direct effect”
Q7b. When a flock of geese flies south in a V-formation, and a storm appears, the storm will slow down the flock of flying geese because the wind blows against the birds, but the V is still maintained. Please explain why the V-pattern is still maintained?		
Key: ● The V-pattern is maintained not due to stronger geese who control the direction of the entire flock // ● but simply by individual goose seeking a place to fly that requires the least effort, which is somewhat diagonally behind another bird.// ● Although the V-pattern is maintained and is heading towards one direction, at times individual goose can fly in various different directions //as it adjust its position to avoid wind resistance and find the sweet spot for flying behind another bird.		
General Rubric: ● Reasons are “No casual agent”, “no alignment”, and “local goal of using least effort against wind resistance”. ● Students could earn one point if their responses indicate any of the above reasons. ● Students could earn a maximum of 3 points for this question.		
Sample Response	Score	Notes
The V-shape is still maintained because of the dynamics of the formation./ All of the geese are exposed in the same way just in different positions (further back, closer to the front)./ Therefore, all of the birds get hit with the same resistance from the storm allowing them to keep their formation intact./	2	● “different positions” indicates “no alignment” + 1 ● “Same resistance from the storm” indicates “same set of interactions” +1
Even with a storm, the V-pattern is optimal under these conditions./ They will still conserve energy./ The aerodynamic formation will help them maintain their formation./ Also, they will have more resistance against the wind./ Their movements will make their flight more smooth./	1	● “conserve energy”, “aerodynamic formation”, “more resistance against the wind” all indicate individual goose seeking a place to fly that requires the least effort – “local goal” + 1
The wind from the storm may slow down the geese, but it doesn't stop them from continuing to follow the "leader"./ This formation helps combat the wind resistance.	0	● Misconception: “leader” – “controlling agent”

APPENDIX H

MISCONCEPTION CODEBOOK

Misconception	Description	Relevant PAIR-C Sequential Features
MC1: Intra-generational change	Creatures change by active adaptation. These changes are passed on to their offspring within their lifetime. Trait change occurs within a generation	F5. Incremental pattern: a. initial pattern is added incrementally
e.g. Industrial pollution will cause the adaptation of darker skin colors to evolve during the lifespan of the moth offspring.		
MC2: Use and Disuse	Changes to individual organisms occur as creatures use particular features more or less. Traits are gained through use or disuse - used traits remain, and those that are not used are lost.	F3. Serial Order F7a. Causal agent: v. direct affect
e.g. A mouse first keeps using its claws to develop stronger claws and then reproduces to pass on the stronger claws to its offspring. A mother giraffe keeps using and stretching its neck to reach higher leaves so that it can directly affect its offspring to have long-neck traits.		
MC3: Directed Variation	Mutations occur as a response to environmental challenges (adaptive mutation) or selection pressure and therefore are always beneficial.	F2. Restricted: a. Uni-directional F7b. Align: iii. Same direction
e.g. Research has shown that white-coated animals have a higher probability to survive and reproduce under heated weather. Therefore, the increased temperature causes white sheep to only mate with white sheep. The pattern of white sheep becoming more common is because the number of white sheep always increases.		
MC4: Teleological conceptions	Changes arise that are purposeful and goal-oriented. Evolutionary forces occur in response to a particular need or predetermined plan.	F7a. Causal agent: iii. Intentional, iv.Teleological
e.g. The moth would have needed to grow darker skin for survival.		
MC5: Anthropomorphism	Assigning human-like conscious intent to the objects of natural selection or to the process itself. Changes are the result of purposeful and goal-directed action provoked by maladaptation.Organisms are characterized by human consciousness, actions, tendencies, emotions, and so on. For example, fighting, hiding, learning, and talking.	F2. Restricted: a. Uni-directional F7a. Single causal agent ii. Special status iii. Intentional iv.Teleological
e.g. White sheep want to mate with white sheep and refuse to mate with brown sheep to avoid having brown sheep offspring.		

<table>
<tr><td>MC6:
Essentialist conceptions</td><td>a species’ outward appearance and behavior are determined by a kind of hidden causal power or “essence”.</td><td>F3. Serial order
F4. Dependent
F7a. Single causal agent</td></tr>
<tr><td colspan="3">e.g. Whether a long-necked giraffe mate with a short-necked giraffe depends on prior mating interaction behaviors.
Mother nature causes moths to change their color from white to black.</td></tr>
<tr><td>MC7:
Transformational views</td><td>An entire population transforms and changes traits as a whole as it adapts.</td><td>F5. Incremental pattern
F6. Cumulative summing</td></tr>
<tr><td colspan="3">e.g. In the next generation, the entire sheep population will change from black fur to white fur as sheep adapt to the hot weather.</td></tr>
<tr><td>MC8:
Events vs. Processes</td><td>a general miscategorization of evolution as a complex event, and not as a process or equilibration.</td><td>F5. Incremental pattern
F6. Cumulative summing
F7b. Align</td></tr>
<tr><td colspan="3">Misconstruing Natural Selection as an Event may contribute to <i>Transformationist thinking</i> (MC7) as adaptive changes are thought to occur in the entire population simultaneously. It can also lead to incorrect “saltationist” assumptions in which complex adaptive features are imagined to appear suddenly in a single generation (MC1).
If it is hard to determine whether the misconception is MC1 or MC7, we can code it as MC8 as it is more general than the others.</td></tr>
<tr><td>MC9:
Absolutes vs. Probabilities</td><td>a general misconception of viewing natural selection as being “all or nothing,” with all unfit individuals dying and all fit individuals surviving.</td><td>F1. Different set of interactions
F7b. Align:
i. Movement: stop-stop, continuous-continuous
ii. proportional:
iii. Same direction</td></tr>
<tr><td colspan="3">e.g. Only white sheep can survive, and all brown sheep will die.
White sheep will stop mating with brown sheep as they dominate the overall pattern
White sheep are always added to the population to create a pattern of white sheep becoming more common.</td></tr>
<tr><td>Undecided</td><td colspan="2">For responses that are vague or statements we aren’t sure how to characterize them into different misconception categories</td></tr>
</table>

APPENDIX I

PAIR-C FEATURES CODEBOOK

Sequential Features	Indicators	Emergent Features	Indicators
1. Distinct Set	Only the fit organisms can survive and reproduce	1. Same Set	All organisms can survive and reproduce
2. Restricted	The fit organisms will only reproduce with other fit ones and cannot be eaten by predators.	2. Random	Any organism can reproduce with anyone else and be eaten by the predators.
3. Serial Order	An organism will reproduce with fit ones first and then with others.	3. Simultaneous	Organism interactions do not follow any temporal constraints.
4. Dependent	Organisms behavior is dependent on other organisms' interactions	4. Independent	Organism interactions do not depend on prior interactions
5. Incremental Change	Organisms change traits by active adaptation within a generation	5. Converging Change	All organism interactions approximate the final pattern over time.
6. Cumulative Summing	Numbers of organisms increase/decrease between time	6. Collective Summing	Proportional change within time, not the absolute number
7a. Causal Agent	"Needing to develop a trait", "allowing", "one species dominate another" are indicators	7b. No Single Causal Agent	No single organism could control others' behaviors intentionally
7b. Align	Whole species changes at once, most fit organism stays most common.	7b. No Align	Variation exists at the end of the process; Individuals with less favorable traits continue to exist.

APPENDIX J

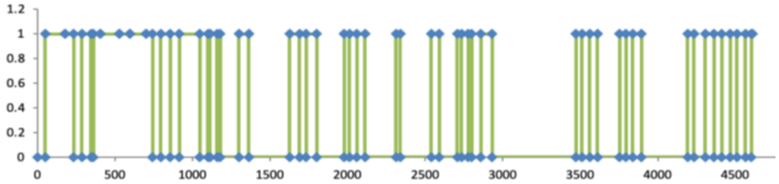
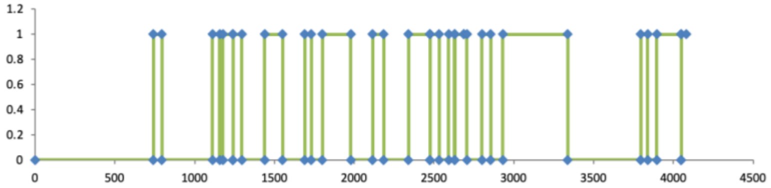
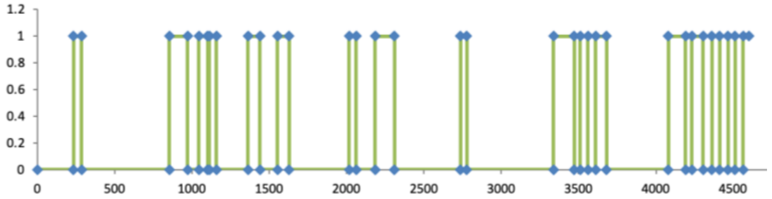
STUDENT ENGAGEMENT CODEBOOK

Indicator	Definition	Learning Behavior	
		Prompt Question Response	Simulation Module Interaction
Passive	Learners receiving information from instructional materials without any overt action	● Not sure how to respond ● Not sure what the question means ● Response too short, such as "all of it"	● Watch simulation videos ● Read texts on PAIR-C features or Darwinian Principles
Active	Learners exhibit visible behavior or physical manipulation	● Describe the simulation environment ● Describe the changes based on observations ● Paraphrase the instructional texts	● Pause, or replay simulation video ● Adjust simulation parameters and observe changes ● Answer prompt questions actively
Constructive	Learners generate or produce additional outputs beyond what is provided in the learning materials	● Predict changes based on initial condition ● Explain why changes happen ● Reflect on the instructional materials ● Elaborate the observation using PAIR-C features or Darwinian Principles	● Ask questions after reading texts ● Check predictions and explain discrepancies ● Compare the recorded simulation video with the actual simulation model ● Answer prompt questions constructively
Off-task	Learners were not engaged in the learning activity	● No response	● No action

APPENDIX K

V-NOTE HISTOGRAM

Comparative Analysis of Student ICAP Engagement Histograms

Name	Student ICAP Engagement Histograms
Eric	Passive:  Active:  Constructive:  Eric spent 1 hour and 17 minutes (approximately 4,500s) finishing the first Peppered Moths simulation scenario. During the learning process, he demonstrated ● 31 instances of passive learning behaviors (e.g., watching tutorials and instructional videos, reading instructional texts, observing data table, etc.); ● 18 instances of active learning behaviors (e.g., adjusting simulation parameters and describing changes observed, go back to the simulation model and check for understanding, take screenshots or input data, etc.); ● 17 instances of constructive learning behaviors (e.g., asking questions after reading texts, comparing simulation contents shown in the instructional video with changes in the simulation model, predicting patterns based on understanding, reflecting on the

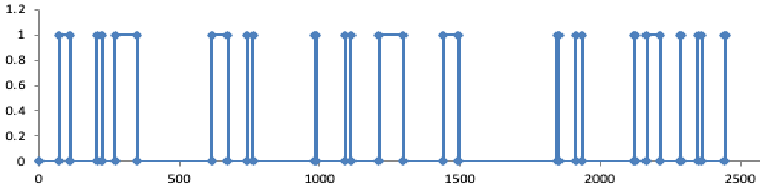
instructional texts and identifying the most important part, explaining reasons for changes, etc.)

- Overall, he spent approximately 29 minutes (37%) on passive activities, 27 minutes (35%) on active activities, and 20 minutes (26%) on constructive activities. Only 2 minutes (3%) on off-task activity.

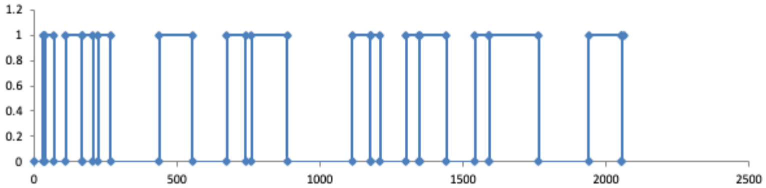
Note that Eric started recording from the preparation phase which took him about 11 minutes to set up the online learning environment before the actual learning activities.

Justin

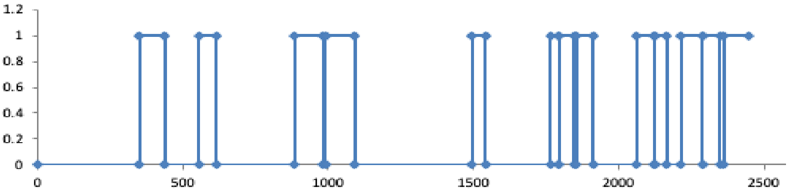
Passive:



Active:



Constructive:



Justin spent 41 minutes (approximately 2,500s) finishing the first simulation scenario. During the learning process, he demonstrated

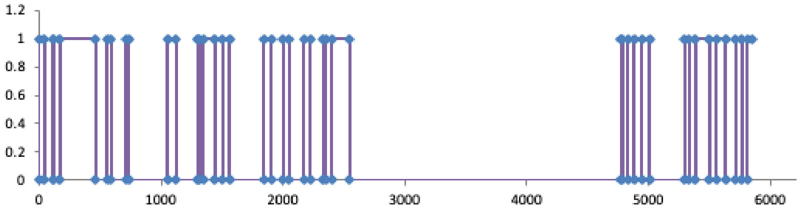
- 16 instances of passive learning behaviors (e.g., watching tutorials and instructional videos, reading instructional texts, observing data table, etc.);
- 16 instances of active learning behaviors (e.g., adjusting simulation parameters and describing changes observed, selecting the correct graph, using simulation to collect and input data, etc.);

- 13 instances of constructive learning behaviors (e.g., providing reflections and explanations based on understanding, etc.)
- Overall, he spent approximately 8 minutes (20%) on passive activities, 17 minutes (43%) on active activities, and 14 minutes (35%) on constructive activities. Only 30 seconds (1%) on off-task activity.

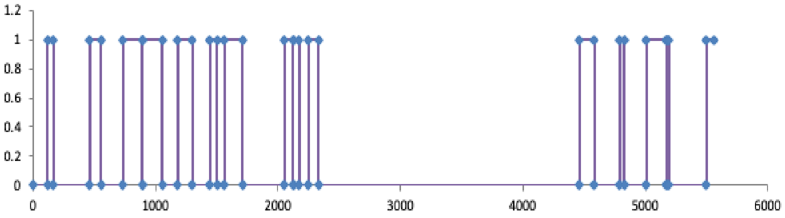
Note that this student showed a tendency for skipping the instructional material, especially the instructional text on PAIR-C features during the learning process. For instance, he spent 3 seconds reading the instructional text about Feature 7a (no direct effect), 2 seconds reading text about Feature 6 (collective summing), and 4 seconds on Feature 7b (not align).

Alice

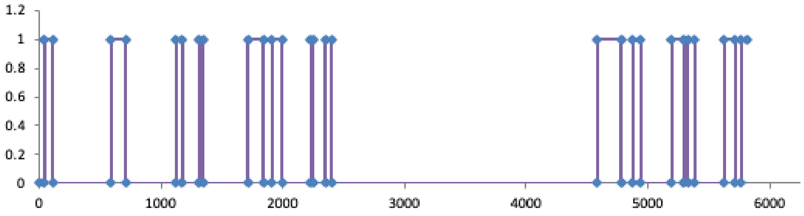
Passive:



Active:



Constructive:



Alice spent approximately 1 hour and 37 minutes (less than 6,000s) finishing the first simulation scenario. During the learning process, she demonstrated

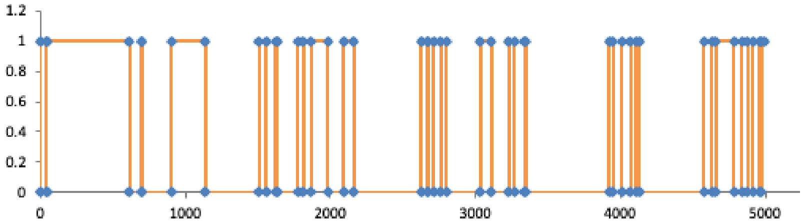
- 22 instances of passive learning behaviors (e.g., watching tutorials and instructional videos, reading instructional texts, observing data table, etc.);

- 15 instances of active learning behaviors (e.g., checking the simulation from time to time while describing the changes; running the simulation and selecting the correct graph, replaying the simulation video while describing changes, etc.);
- 14 instances of constructive learning behaviors (e.g., asking questions, explaining the reasons for the changes while adjusting the simulation, reflecting on the possible answer and comparing it with the previous answer, etc.)
- Overall, she spent approximately 23 minutes (24%) on passive activities, 23 minutes (24%) on active activities, and 18 minutes (19%) on constructive activities.

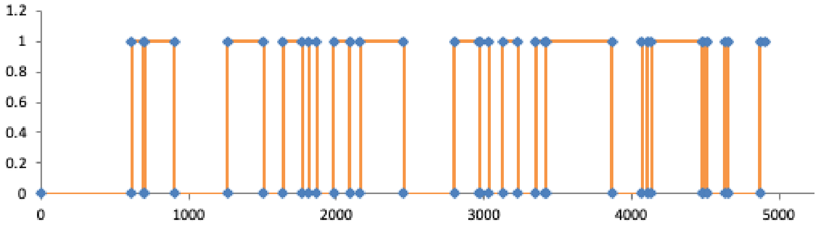
Note that the student was off-task for approximately 31 minutes after 40 minutes of engaged learning and then returned to finish the remaining module at time point 01:14:00 (2,400s). The student seemed to take a break and continued with a similar level of engagement after the break.

Evelyn

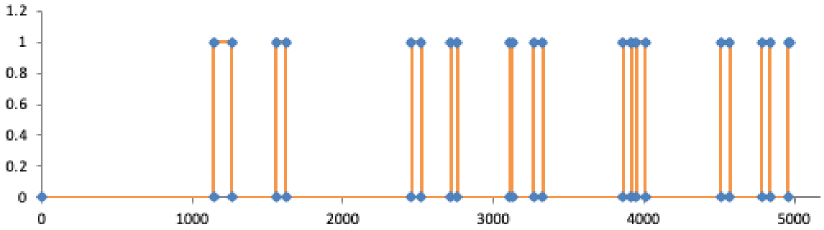
Passive:



Active:



Constructive:



Evelyn spent 1 hour and 23 minutes (approximately 5,000s) finishing the first simulation scenario. During the learning process, she demonstrated

- 23 instances of passive learning behaviors (e.g., watching tutorials and instructional videos, reading instructional texts, observing data table, etc.);
- 17 instances of active learning behaviors (e.g., randomly manipulating simulation parameters, slowing down the simulation video speed and describing observations, using simulation to collect and input data, copy and pasting instructional text, etc.);
- 11 instances of constructive learning behaviors (e.g., providing reflections and explanations based on understanding, etc.)
- Overall, she spent approximately 29 minutes (35%) on passive activities, 40 minutes (49%) on active activities, and 10 minutes (13%) on constructive activities. Only about 2 minutes (2%) on off-task activity.

Note that this student did not show any off-task activities during the simulation module learning process.

In each histogram, the x-axis represents time duration in seconds while the y-axis indicates the number of events. The blue node represents each data point shown in the event track. When the nodes connect and formulate a bar, it indicates an independent observation.

APPENDIX L

LEARNING MODULES

Links:

1. Technology in Action Part I (Study Overview and Consent)
2. Technology in Action Part II (Pretest and Survey)
3. Technology in Action Part III & Part IV:
    a. PAIR-C ABM Intervention Group:
        i. The PAIR-C Peppered Moths Simulation Module (Part III)
        ii. The PAIR-C Rock Pocket Mice Simulation Module (Part IV)
    b. Regular ABM Control Group:
        i. The Regular Peppered Moths Simulation Module (Part III)
        ii. The Regular Rock Pocket Mice Simulation Module (Part IV)

APPENDIX M

IRB APPROVAL

ASU Knowledge Enterprise Development

EXEMPTION GRANTED

Michelene Chi
Division of Educational Leadership and Innovation - Tempe
480/727-0041
Michelene.Chi@asu.edu

Dear Michelene Chi:

On 4/7/2022 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Facilitating Students’ Understanding of Emergent Causal Mechanisms in Natural Selection Using Agent-Based Models
Investigator:	Michelene Chi
IRB ID:	STUDY00015726
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• ASU Student Recruitment/Data Collection, Category: Other. • Consent Form, Category: Consent Form; • IRB Application Form, Category: IRB Protocol; • Recruitment Flyer, Category: Recruitment Materials; • Supporting Document, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 4/7/2022.

In conducting this protocol, you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

REMINDER - - Effective January 12, 2022, in-person interactions with human subjects require adherence to all current policies for ASU faculty, staff, students and visitors. Up-to-date information regarding ASU's COVID-19 Management Strategy can be found here. IRB approval is related to the research activity involving human subjects, all other protocols related to COVID-19 management including face coverings, health checks, facility access, etc. are governed by current ASU policy.

Sincerely,

IRB Administrator

cc: Man Su